

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:39:32 ; Search time 5957.55 Seconds
(without alignments)
7094.489 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccacgcgtccggcttctgt.....tgaggactgagngccaagtg 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_om.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vi.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1459.2	57.0	2421	10	AF060173	Rattus no AF060173
2	1118	43.6	2104	9	HSMB02697	AL359592 Homo sapi
3	472.6	18.4	212183	2	AC087893	AC087893 Homo sapi
C 4	211.4	8.3	183481	2	AC011596	AC011596 Homo sapi
C 5	91	3.6	11856	1	AE007266	AE007266 Sinorhizo
6	89.4	3.5	148432	3	AC004642	AC004642 Drosophil
7	89.4	3.5	157851	2	AC020509	AC020509 Drosophil
C 8	89.4	3.5	300542	3	AE003462	AE003462 Drosophil
9	89	3.5	198677	1	AE001863	AE001863 Deinococc
C 10	84	3.3	19834	3	CEK01F9	222175 Caenorhabdi
C 11	82.8	3.2	10344	1	AE004767	AE004767 Pseudomon
C 12	81.8	3.2	62204	2	AC067757	AC067757 Homo sapi
C 13	77	3.0	11286	1	AE008141	AE008141 Agrobacte
14	74.2	2.9	81020	8	AB026645	AB026645 Arabidops
15	74.2	2.9	81020	24	AC024128	AC024128 Arabidops
16	70	2.7	13462	1	AF218267	AF218267 Pseudomon
17	69.6	2.7	2622	10	AF060174	AF060174 Rattus no
18	69.2	2.7	1638	9	AF052188	AF052188 Homo sapi
19	69.2	2.7	4948	9	AB018278	AB018278 Homo sapi
20	66.2	2.6	2315	5	DYGSV2TRAN	L23403 Discopysge o
21	66.2	2.6	3190	9	BC000776	BC000776 Homo sapi
22	66.2	2.6	4353	9	AB018279	AB018279 Homo sapi
23	66	2.6	3660	10	RATSV2BA	L10362 Rattus norv
24	64	2.5	35000	1	AB000617	AB000617 Bacillus
25	64	2.5	22160	1	BSUB0002	Z99105 Bacillus su
26	63	2.5	2268	6	AX079057	AX079057 Sequence
27	62.4	2.4	10029	1	AE007908	AE007908 Agrobacte
28	61.6	2.4	16775	1	AB000735	AB000735 Nocardiol
C 29	61.6	2.4	40699	3	CEZK637	Z11115 Caenorhabdi
30	59.8	2.3	7424	1	AB050935	AB050935 Pseudomon
C 31	59.2	2.3	11886	1	AE004638	AE004638 Pseudomon
32	58	2.3	159167	2	DMBR42L18	AL122031 Drosophil
33	57.2	2.2	37221	2	AC012886	AC012886 Drosophil
34	57.2	2.2	126340	2	DMBR18C8	AL122030 Drosophil
C 35	57.2	2.2	299537	3	AE003438	AE003438 Drosophil
36	56.6	2.2	3868	4	S47919	S47919 p87-transpo
37	56.6	2.2	7080	1	AF305325	AF305325 Comamonas
38	56.4	2.2	35654	1	SC7H1	AL021411 Streptomy
39	53.8	2.1	17641	1	AE001804	AE001804 Thermotog
40	53.2	2.1	1885	6	A48158	A48158 Sequence 5
41	51.4	2.0	297	9	H518A12R	Z65016 H.sapiens C
42	51	2.0	134092	2	H0702G05	AL442105 Oryza sat
43	50.8	2.0	3663	10	RATSV2A	L01788 Rattus norv
44	50.8	2.0	3844	10	RATSV2AA	L05435 Rattus norv
45	50.2	2.0	2132	10	RNOCTK	X98334 R.norvegicu

ALIGNMENTS

RESULT 1	AF060173	AF060173	2421 bp	mRNA	ROD	21-NOV-1998
AF060173	LOCUS	Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.				
DEFINITION	AF060173					
ACCESSION	AF060173.1	GI:3901267				
VERSION						
KEYWORDS						
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE	1 (bases 1 to 2421)					
AUTHORS	Janz, R., Hofmann, K. and Sudhof, T.C.					
TITLE	SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles					
JOURNAL	J. Neurosci. 18 (22), 9269-9281 (1998)					
MEDLINE	99019745					
REFERENCE	2 (bases 1 to 2421)					
AUTHORS	Janz, R., Hofmann, K. and Sudhof, T.C.					

TITLE Direct Submission
JOURNAL Submitted (16-Apr-1998) Center for Basic Neuroscience, Department of Molecular Genetics, HHMI, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA

FEATURES
Source Location/Qualifiers
1..2421
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
1..2421
/gene="SVOP"
301..1947
/gene="SVOP"
/note="synaptic vesicle protein; similar to transport proteins"
/codon_start=1
/product="SV2 related protein"
/protein_id="AAC78627.1"
/db_xref="GI:3901268"
/translation="MEEDLQRLQPLVVKFRRTGESARSEDAAAGEHUUVQIEGVKRVG
LEAVLDGAAAPKPEFANPTDDTFMVEDAVEAIGFRQWKLSDVLTGLAWMADAMEM
ILSILAPOLHCERLPSQVALLTSVVFIMGSSSTLVGNISDOYGRKTLGISVFWT
LYYGILSAFAPVYSWILVLGLVFGIGVPOSVTLYAEEFLPMKARAKILLIEFWA
IGTVREVLAVFMSLGRWILLISAPRLVFAVLCPLWPESARYDVLGNQEKIA
TLKRIATEANGAPMLPKLIISQEDRGRNRDLFTPHFRWTLILLNFIWFSNFSYGL
VLLTTELFOAGDVCISIRKKKAVEAKCSLACEYLSKEDYMDLLWTLTSEFFQVAVYTLW
VIDRLGRKTMALCFVIFSLCSLLILFICIGRNVLTLLLFIARAFISGGFOAAVYVTP
VYPTATRALGLGTCGMARVAGALITPFIQVLMLESSVYLTLAVISGCCLLAALASCEL
PIETKRALQESSHREWGEMVGRGNTGVPKRSNGSOE"

BASE COUNT 459 a 658 c 735 g 569 t

ORIGIN

Query Match 57.0%; Score 1459.2; DB 10; Length 2421;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 1825; Conservative 0; Mismatches 343; Indels 30; Gaps 13;

QY 41 ggtccctcccaacagctgctgccaggaggaagcgcggtgctgtccagctt 100
DB 168 GGCTCCCTCCCCAGCAGCTGC---CCAGCGAGGAAGCCGC---GGAGCTGTCCAGCAT 220
QY 101 ccggtgtgaaacccgaggggtcgtcaccaccactaccatgtagggccatgagaag 160
DB 221 CCCGGTGCTCAAAAGCCGAGCGCGCCCTGCTTC---ACCATGTAAGGTGCATGCCCA- 276
QY 161 ggtcatctctggcagcgagcgagacatgagaggaacttattccagctaaaggagctcgg 220
DB 277 GGCTCGCCCTGGCGCAGCGCGGACATGGAGGAGGACCTGTTCAGCTCAGGCGAGTTCGCCG 336
QY 221 gtttgaattccgtcgcacagcgagagtgcaaggtcagagagcagcagcggttcaggaa 280
DB 337 GTGGTGAATTCGGCGCACAGGAGAGAGCGACGGTCCAGAGGACGCGGCTTCGCCGG 396
QY 281 gagcatgaagtccagatgaaaggggtccacgctgggacctagagcgtggaagctggatgat 340
DB 397 GAACATGATGTTTCAGATTGAGGGGGTCCGAGTGGGCGCTAGAAAGCTGTCCGAGCTGGATGAT 456
QY 341 ggggcagctgtgccaaaggagtttgccaatccacacgatgatactttcatggttggaagat 400
DB 457 GGACGACTGTCCCAAGAGAGTTGGCAATCCACCATGATGACACTTTCATGGTTCGAAGAT 516
QY 401 gcadgaaagccattggctttgaaaaatttcagtgggaagctgtgtgtctcaactggcttg 460
DB 517 CGGTGGAAAGCCATTGGGTTTCGGAAGATTCCAGTGGGAAGCTCTGTCTTCACCGGGCTTG 576
QY 461 gcttggatgctatgcatggagatgatctcagcactctcctggcaccacagctgcat 520
DB 577 GCTTGGATGGCGGACGCGCATGGAGATGATGATCTGAGCATCTTCGGCGCCTCAGCTGCAC 636
QY 521 tgcagtgagggtcccaaggcaggtggcattgtgcacctcgggtggtctttaggc 580
DB 637 TGCAGTGGGCGACTCCCGCAGGCTGGCAGGTGGCGCTGTGACTTCGGTGGTCTTCATTGGT 696

QY 581 atgagtccagctccacgctctggtgggaaatatactcagaccagatagcgagcaggaacagagg 640
DB 697 ATGATGTCCAGTTCTACGCTCTGGGAAACATCTCGATCAGATATGCGAGAAACAGGG 756
QY 641 ctgaagatcagcgtgctgtggactctgtactatgcatccttagtgcatttgcgcctg 700
DB 757 CTGAAGATCAGTGTGTCTGGACCCCTGTACTACGGGATCCTCAGCGCTTTCGCGCCAGTG 816
QY 701 tatagctggtatcctggtccggggcctgggtgggcttcggggtcggagagttccccag 760
DB 817 TATAGCTGATCTGTGTCTCCGAGGCCCTCGTGGGCTTTGGGATGGAGGGTGCCTCAG 876
QY 761 tcggtgacgctgtatcccgagttccttcccatgaaagccagagcagataatgtatttgcgtg 820
DB 877 TCGGTGACCTGTACGCCGAGTTCTCCCATGAAGCCGACAGGCAAGTGCATTTTGTGCTG 936
QY 821 attgaggtattctgggccatcgggacagtggtcgaagtcgtcctggtgctgtgtcgtgagt 880
DB 937 ATTGAGGTTTTCGGGCCATCGGACCGTTCGAGGTGCTTCTGGCTGTGTGTGTGTGATG 996
QY 881 ccagcgtggctggcgttgggtgctcactctcagctgtcccgctctctctcttcttgc 940
DB 997 CCCAGCTTGGCTGGCGTGGCTGCTGCTGCTGCTGCGCGCTCCACTACTGTCTTTGCT 1056
QY 941 gtgctgtttctgctgctgaaagtcaaggtatgattgctgtcaggaagaccagga 1000
DB 1057 GTTCTGTGTTCTGGCTGGCAGAGAGTGTAGTACATGTGCTGTGCTGGAAACACAGGAA 1116
QY 1001 aagcgaatcgccaccttaagagatagcaactgaaacgagagctcccatgccgtgggg 1060
DB 1117 AAGCGATTGCTACCTTAAAGCGGATCGCAACAGAAATGAGAGCCCGCTCTGGGG 1176
QY 1061 aaactcatctccagcaggaagcagcaggaacaaatgagggaccttttcaacacccat 1120
DB 1177 AAGCTCATCATCTCCACAGAGGAAGCCGAGGCAAAATGAGGGACCTTTTCACACCCAC 1236
QY 1121 tttagatgacaacttctgctgtgtgtttatatgtgtttcccaatgcattctcttactac 1180
DB 1237 TTTCTGTGGACAACCTGCTGCTGTGTTATATGTGTCTCCAATGCTTCTCTATTAC 1296
QY 1181 ggtttagttctactaccacagaaactctccagcgggagatgctgcggcagctccag 1240
DB 1297 GGCTGTGTTCTGTCTACACAGAACTCTTCAGCGCGGAGATGTTTCAGCATCTCCAGC 1356
QY 1241 cgaaagagctgtagagcgaatgcagcctggcctgcagtaacctgagtggaggagat 1300
DB 1357 CGGAAGAAGCGGTGGNAGCCAAATGCAGCTGGCTGTGTGAGTACCTCASCAGAAGAGAT 1416
QY 1301 tacatggaactgtgtgagaccacctctctctgagtttccagggtgctctgtgactctgtg 1360
DB 1417 TACATGACCTGCTGTGGACACGCTGTCTGAGTTCCAGGTGTCTCTGTGACTCTGTGG 1476
QY 1361 attattgacgctggggcgcaagaagaccatggcctgtgcttctgtcatctctctctc 1420
DB 1477 GTCATGACCGCTTGGCGCCCAAGAACCATGGCTCTGTGTGTGCTATCTTTTCCTC 1536
QY 1421 tgcagcctctgctgtttatctgtgttgggaagaaatgctcactctgttactcttcaat 1480
DB 1537 TGCAGCTCTCTGCTGTTCATCTGCATTGGAAAGAAATGTGCTAAACCTCTTTACTGTTCATT 1596
QY 1481 gcaagagcgtttattctggagccttcaagcggcatatatttacaacctgaggtctac 1540
DB 1597 GCAAGACGTTTTATTCTTGAGGCTTCCAAAGCAGCCCTACGTTTACACGCTGAGGTGAT 1656
QY 1541 ccacaggaacgcgggacctcgccctgggcacctgcagcggcctggaagagtggtgct 1600
DB 1657 CCAACGGCAGAGGGCGCTGGGCCCTGGGCACCTGCAGCGGCATGGCGAGAGTGGCGCG 1716
QY 1601 ctcatcactccgttcatcgcccgaggtgtagtctggaatcctctgtacctgactctgca 1660
DB 1717 CTCATCACTCCATTCATAGCTCAGGTGATCTGGAATTTCTCCGTGTACCTGACCTGGCC 1776
QY 1661 gttacagtggtgctgctcctcctggctgcctcctgctgtttttgcccattgagacc 1720

1777	GTCTACAGTGGCTGCTCTCTTGTGCTGCTTGCCCTTCCTGCTTCTGCCATCGAGACC	1836	gene	/dev_stage="adult"
	1. .437			/tissue_type="amygdala"
1721	aaaggcgagactcaggaatccagccaccggagtgggccagagatggctcgccga	1780	CDS	/gene="DKFZp761H039"
	<1. .437			<1. .437
1837	AAAGGCCGAGCACTGCAGAGTCCAGCACCGGAGTGGGGCCAGGAGATGTTGGCCGA	1896		/gene="DKFZp761H039"
				/note="similarity to SV2 related protein (Rattus norvegicus)"
1781	ggaatgcacggtgcaggtgtttaccagggtcgaaactctggctctcaggaaatagtagaccgatg	1840		/codon_start=3
	111			

```
OY 2115 aatgccagcaaccaatcagacacacccagagtcacccggcccaaacccctcagtgaacaac 2174
|||||
Db 721 AATGCCAGCAACCAATCAGACACACCCAGAGTCACCCGGCCAAACCCCTCAGTGAACAAC 780
|||||
OY 2175 caaataatctctgtatagatcacgttcagagctcagggcccatgtgacacctcgtgtccacc 2234
|||||
Db 781 CAAAATATCTCTGTGTAGATACCGTCCAGGCTCAGGCCCATGTGACACCCGCTGTCCACC 840
|||||
OY 2235 caccgagctgttcagtagtttctccacacacacagcccccaggctttcttcttcttgaaa 2294
|||||
Db 841 CACCGGACCTGTTTCAGTAGTCTTCTCCACACCCACAGCCCCAGGCTTCTTCTTCTTGA 900
|||||
OY 2295 ttccagggcagctaggtgtgtgtcagcagcatttctcctggcagggggcccccgggttgc 2354
|||||
Db 901 TTCCAGGCGATCTAGGTGTGTGTGAGCAGCTATTTCTGGCAGGGGCCCCCGGTTTC 960
|||||
OY 2355 ctccctagaacctgacagtggaattctctgacagatgagacatgtgcatccaacagggag 2414
|||||
Db 961 CTCCTAGAGCCCTGACAGGTGGATTCTCTGGCAGATGGACATGGTCATTCAAACTGGAG 1020
|||||
OY 2415 ccacatgcccccacccagccctntlggagttgcccgtttgtggcaccagagatccaga 2474
|||||
Db 1021 CCACATGCTCCACCCAGCCCTCTGGAGTGTG-CCGTTGTGGCACCAGAGATCCAGA 1079
|||||
OY 2475 ttgtcctcgtgggacagctgggtttcaccacaggtgacacaccccaaacgcgttaccccc 2534
|||||
Db 1080 TGTGTCCCTGGGACAGCTGCTGTTCACCCAGGTGACAACTGACAGACGCGCTTACCCCC 1139
|||||
OY 2535 tggggaactgaggaactgagagccaagtg 2562
|||||
Db 1140 T-GGAAC-CAGGACTGAGGCCAAAGTG 1166
|||||
```

```
RESULT 3
AC087893 LOCUS
DEFINITION Homo sapiens chromosome 12q clone RP11-423G4, WORKING DRAFT
AC087893
VERSION AC087893.7 GI:14328965
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 212183)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Huiy,K.S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulisege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
```

```
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Slisson,I., Sodergren,E., Sonaie,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 212183)
Worley,K.C.
Direct Submission
Submitted (06-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 8, 2001 this sequence version replaced gi:13928634.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCHH
Center clone name: RP11-423G4
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 201526 bases at least Q40
Consensus quality: 205059 bases at least Q30
Consensus quality: 206774 bases at least Q20
Estimated insert size: 205251; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 5.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 41137: contig of 41137 bp in length
* 41138 41237: gap of unknown length
* 41238 69033: contig of 27796 bp in length
* 69034 69133: gap of unknown length
* 69134 89390: contig of 20257 bp in length
* 89391 89490: gap of unknown length
* 89491 100180: contig of 10690 bp in length
* 100181 100280: gap of unknown length
* 100281 112787: contig of 12507 bp in length
* 112788 112887: gap of unknown length
* 112888 127738: contig of 14851 bp in length
* 127739 127838: gap of unknown length
* 127839 137576: contig of 9738 bp in length
* 137577 137676: gap of unknown length
* 137677 145046: contig of 7370 bp in length
* 145047 145146: gap of unknown length
* 145147 154832: contig of 9686 bp in length
* 154833 154932: gap of unknown length
* 154933 164690: contig of 9758 bp in length
* 164691 164790: gap of unknown length
* 164791 172212: contig of 7422 bp in length
* 172213 172312: gap of unknown length
* 172313 179273: contig of 6961 bp in length
```


VERSION
SOURCE
ORGANISM

REFERENCE
AUTHORS

FEATURES
source

BASE COUNT 55581 a 50102 c 50442 g 54260 t 1798 others
ORIGIN

Query Match 18.4% Score 472.6; DB 2; Length 212183;
Best Local Similarity 90.7% Pred. No. 1.2e-112;
Matches 525; Conservative 0; Mismatches 51; Indels 3; Gaps 2;

Qy 1586 gcaagagtggtgtctctcatcactccgttcacgtcccccaggtgatgctggaatcctctgtg 1645
Db 68423 GAAGGGATGCTTTCACCGACCCGCTGTCTCCCTAGGTCATGCTGGAATCTCTGTG 68482

Qy 1646 tacctgaacttggaattacagtggtctgcctctgcctgcctgcctgcctgcctgcctgtt 1705
Db 68483 TACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTCCCTGGCTGCTCCCTGCTGCTT 68542

Qy 1706 ttgcccattgagaccacaaaggcagagctgcagagtcagcagccacccggagtgagg;cag 1765
Db 68543 TTGCCCATGAGACCAAGCCGAGGACTGCAGGAGTCCACCCACCGGAGTGGGCCAG 68602

Qy 1766 gagatggtcggccgaggaatcacagtgctcaggtgttacaggctgaactctggctctcag 1825
Db 68603 GAGATGCTGGCCGAGGAATCACGGTGCAGGTGTACACAGTCCGAATCTGGCTCTCAG 68662

Qy 1826 gaatgtagccatggggagctgagctggtctcttggagctgcagagcttgggggagctggc 1885
Db 68663 GAATGAGCCATGGGGAGCTGAGCTGCTCTTCAGGCTGCAGAGCTTGGGGGCTGSC 68722

Qy 1886 agggcccaactggggcactgattgtcactgcagacatcaagaactcaccacaaagatgta 1945
Db 68723 AGGCCCAACTGGGCACTGATTGTCTACTGGCGACATCAAGAAGTCAACCAAGATATGA 68782

Qy 1946 cctgacacaaagggtttgtgtctgactcagttgtctcatttcattgaagttccacc 2005
Db 68783 CTGTGACCAACAGGGTTTGTCTGTGACTCAGTTTGTCTATCTTCATTTAGGTCACCC 68842

Qy 2006 agggat-ggggagatgtttgctctcagggggttctctgtatatgtggtgaaagcttgttc 2064
Db 68843 AGGGATGGGGAGATGTTTGTCTCTTANGGGGTTCTCTGNTATGCGGTGAAAGCTCTGTC 68902

Qy 2065 ataacctgtggtatcataggaagaaactaccatattaggaaggtctgttaatgccagca 2124
Db 68903 ATAACTGGGGATCATACATCGGGAGACCAACCCATATTAGACGGTCTGCGCATCGCAATCA 68962

Qy 2125 accaatcacacacccacccagagtcacccggcccaaccct 2163
Db 68963 ACAATTC--AAACCATCAAACTCACCCGGTCAACCCCT 68999

RESULT 4
AC011596/c
LOCUS
DEFINITION
SEQUENCE, 8 unordered pieces.
AC011596

AC011596.22 GI:15029431
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 183481)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,K., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loisleged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Nehtson,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogutu,M., Okwono,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczka,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 183481)
Worley,K.C.
Direct Submission
Submitted (08-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 30, 2001 this sequence version replaced gi:14547725.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

Center project name: HMKK
Center clone name: RP11-117B7
----- Summary Statistics

Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 208349 bases at least Q40
Consensus quality: 213493 bases at least Q30
Consensus quality: 216913 bases at least Q20
Estimated insert size: 193550; sum-of-ctigs estimation
Quality coverage: 9.2x in Q20 bases; sum-of-ctigs estimation

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT


```

similarity to short chain"
/codon_start=1
/transl_table=11
/product="Probable"
/protein_id="AAK65450.1"
/db_xref="GI:14523918"
/transl_table="MOLKSRVIVTGASSGLGAATVPMIAQEGATVLGLDLKPPAGEE
PAELGAARFRNADVTNEADATAALAFKQEFQGHVHCLVNCAGTAPCEKILGRSGPH
ALDSFARTVNLIGTFNEMRLAEVMSQGEPADEGGRVIVNTASTIAAFDQGIQQA
YAASKGVAALTPAARELARFAGIIRVVTTIAPGIDFTPMAGMEQDVODALAASVPFPP
RLGRAEYAAALVKHICENTMLNGEIVRLDGLALMRAP"
3995..5065
/gene="Sma1454"
/EC_number="5.5.1.1"
/3995..5065
/gene="Sma1454"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; global similarity to OruR from
P. aeruginosa (GI: 1657970) which regulates ornithine
utilization; pfam domain similarity to N-terminal
helix-turn-helix motif"
/codon_start=1
/transl_table=11
/product="Putative transcription factor"
/protein_id="AAK65451.1"
/db_xref="GI:14523919"
/transl_table="MNEKQPSHWRMTAPCFMDDTLECLRHRGVDAGPLLAQAGLP
SIVTGPVSANQYAFHVAHQAMDEFFEGEGARPMRAGSPALICHAITLSTALEHALR
RAKRLVLENSPAGELVYENGLAQVLKDGATRSFAFRTFMIILHGVNCLWLGRR
LPTRWDVFCSSPAGDYLRFEGFQDFDQRTLVFDAYELKLPPIRDERALKHFL
RHAPANLLVRYHDAGLSAATQRHLHATAPSAWPGFEATAARMRIAAPTLLRRLRQEG
QTVRSIKEDLLRALAMEALADGRNTNAQLAVELCFSEPSAFHRAFRKWTGSKSPAQFR
NATOTDFAESSSPKRKTQPOES"
complement(5047..5565)
/gene="Sma1455"
/5047..5565
/gene="Sma1455"
/complement(5047..5565)
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65452.1"
/db_xref="GI:14523920"
/transl_table="MRTGTCYRAHDPWAFKPAASGDGAAIKARENPKGVKTYLIAL
SMTAVKEANQGFHRIIDPCVLCSDYIDCADIALDTAERGRAEHGVSLEMACSWAGA
FAECRPPASWAIYNRLHPRCTAGILVPSFAPGTEAGDRNLVLTWPGDLPFLRLDVTDP
SGRLPKDQLSWG"
complement(5562..5951)
/gene="Sma1456"
/5562..5951
/gene="Sma1456"
/complement(5562..5951)
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65453.1"
/db_xref="GI:14523921"
/transl_table="MAQAKIQSDGGPILLSYMDKGGKIATVOQAVFGFMSKTQLAET
AGLARLTLRPKSRGKTQNRLEMLIISRVTDWAGKEQAMAWYRAQPLPAFGGR
TABALVKDGKAAAVRDYLDHMLGGFA"
complement(6059..7063)
/gene="Sma1459"
/6059..7063
/gene="Sma1459"
/complement(6059..7063)
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to
malate-dehydrogenase from M. fervidus (GI: 80950)"
/codon_start=1
/transl_table=11
/product="Putative"
/protein_id="AAK65454.1"
/db_xref="GI:14523922"

```

```

/transl_table="MTPVETSWEEIEAVCLLEALTLHGAAAEARAVAGAITRAADGN
RVCGLYLPFIICRHILAIKGVDEAVPEVTRRGVTVTVDAKSGFAHPAIAAGTALIDL
ARQAGLAAMAVRNSNCLALGHHVRPLADAGLIGICVSNAPASVAPGCFATRALFETNP
LAFAPSKEGAPTIVDQSGMSAVTKTEMLRRDRGEAIPICWAGDQNCQPTTDAATGL
EGSLLPAGRGKGANIALLVLEAALTGSAALTSASAFNEEGPPHVGQFLIALDPD
HFAAGHFSEAMDNLVASHDAAGVRLPGHFGKRPQVCDVADULWKKAVLLSKNRKQKPG
"
complement(7060..8184)
/gene="Sma1461"
/7060..8184
/gene="Sma1461"
/EC_number="5.5.1.1"
/function="Small molecule metabolism; Central intermediary
metabolism; amino acids"
/note="glimmer prediction; global similarity to muconate
cycloisomerase in P. putida (GI: 77758); Pfam domain
similarity to mandelate racemase / muconate lactonizing
enzyme family"
/codon_start=1
/transl_table=11
/product="Putative muconate cycloisomerase"
/protein_id="AAK65455.1"
/db_xref="GI:14523923"
/transl_table="MKNISNVRVPLVPLKQPYHWSYGIRESEFAVNIIEIAUDGTV
GIGECTVAPQDTGTAAITLYRLAKHLVGHSPHDVAPLARIHFOEYILGHGANIMRAANO
IFSGIDMAMMDLQKLAGLPVHOLLGGAHRKAVGYFYFLOQETAELARDAAVGHAOG
ERYFYLVKVGEGEKLDLEITAAVRGEIGDARLDANEGWSHIDAINCRKLEKDYIEF
IEQPTVSWSIIPAMAHVREKVIPTVAIQAAFTLYDVYIEICRQRAADMICIGPREIGGI
OPMKAAVAEAAAGKATCICIHSSFTTITTCAEHHIIGLIAINLDDGNQIMQVLQVEDIV
SSPDLTPKNGLDLAFKPGIGLQALAEPLVAEGEGRYAASR"
complement(8311..9201)
/gene="Sma1462"
/8311..9201
/gene="Sma1462"
/function="Cell processes; Transport of small molecules"
/note="glimmer prediction; global similarity to
choline-binding protein precursor from B. subtilis
(GI:7404430) and several putative ABC transporter
proteins; adjacent ORFs encode other putative members of
the choline transport system"
Query Match 3.6%; Score 91; DB 1; Length 11856;
Best Local Similarity 48.3%; Pred No. 8.9e-13;
Matches 324; Conservative 0; Mismatches 335; Indels 12; Gaps 2;

Qy 383 acttctatggtggaagatgcagtggaagcattggctgttggaataatttcadgtgaagctg 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1423 ACAATCAGTAGACGATCGCTCGATCGCGCGGACGGGAACCTTACCAGCGCAGACTG 1364

Qy 443 tctgttctactggttggcttgatggtcgcacatgcagatgcagatgcagctc 502
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1363 ATGGCCATATTCGGTCTGGTGGCTGGCGATGCCATGCCAGGTGCTGGCGTTC 1304

Qy 503 ctggcaccacagctgcattgcagtgaggtctccacagctggcaggtggcattgctgacc 562
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1303 ACCGCCCATCGATCGCGCGCCACTTTTCGGCTGACATGTCGCGAGGCACTTCACACGGGC 1244

Qy 563 tcggtggtctttagcatgatgtccagctccacgctctgggaaaaatactcagaccag 622
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1243 ACGTTTTCTTCTCGGAATGCTGTTCCGGCGCGCGGTTTCGGCAGGCTCGCGGACCGG 1184

Qy 623 tacggcagaaaaacagggctgaagatcagcgtgctgaggactctgtactatggcatt 682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 ATCGCGCGCGCGGCTGCTGATCGCCACCGTCCGCTGCCATGCCGCTTCCTGCTGCTG 1124

Qy 683 agtgcatttgcgcctgtatagctggatccttgctccggggcctggtggcctcg 742
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 TCGCTCTTTCGCGCAGGACTTTACCGTTTTGCTTCTGCTGGGTTTCTCAGCGGCGCAGCT 1064

Qy 743 atcgaggagg---agtccccagtcggtgacgctgtatgccagttccttccatgaaagcc 799
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 GTCGCGGCGCACACTGCGCGGTTCGACTATCGCATGATGCGGAGTTCTCTCGCGCGCGGAAC 1004

```

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 730: contig of 730 bp in length
* 731 830: gap of 100 bp
* 831 1522: contig of 692 bp in length
* 1523 1622: gap of 100 bp
* 1623 2322: contig of 700 bp in length
* 2323 2422: gap of 100 bp
* 2423 3119: contig of 697 bp in length
* 3120 3219: gap of 100 bp
* 3220 3925: contig of 706 bp in length
* 3926 4025: gap of 100 bp
* 4026 4710: contig of 685 bp in length
* 4711 4810: gap of 100 bp
* 4811 5499: contig of 689 bp in length
* 5500 5599: gap of 100 bp
* 5600 6293: contig of 694 bp in length
* 6294 6393: gap of 100 bp
* 6394 7102: contig of 709 bp in length
* 7103 7202: gap of 100 bp
* 7203 7920: contig of 718 bp in length
* 7921 8020: gap of 100 bp
* 8021 8722: contig of 702 bp in length
* 8723 8822: gap of 100 bp
* 8823 9516: contig of 694 bp in length
* 9517 9616: gap of 100 bp
* 9617 10317: contig of 701 bp in length
* 10318 10417: gap of 100 bp
* 10418 11127: contig of 710 bp in length
* 11128 11227: gap of 100 bp
* 11228 11922: contig of 695 bp in length
* 11923 12022: gap of 100 bp
* 12023 12715: contig of 693 bp in length
* 12716 12815: gap of 100 bp
* 12816 13518: contig of 703 bp in length
* 13519 13618: gap of 100 bp
* 13619 14318: contig of 700 bp in length
* 14319 14418: gap of 100 bp
* 14419 15123: contig of 705 bp in length
* 15124 15223: gap of 100 bp
* 15224 15919: contig of 696 bp in length
* 15920 16019: gap of 100 bp
* 16020 16721: contig of 702 bp in length
* 16722 16821: gap of 100 bp
* 16822 17533: contig of 712 bp in length
* 17534 17633: gap of 100 bp
* 17634 18322: contig of 689 bp in length
* 18323 18422: gap of 100 bp
* 18423 19120: contig of 698 bp in length
* 19121 19220: gap of 100 bp
* 19221 19932: contig of 712 bp in length
* 19933 20032: gap of 100 bp
* 20033 20735: contig of 703 bp in length
* 20736 20835: gap of 100 bp
* 20836 21544: contig of 709 bp in length
* 21545 21644: gap of 100 bp
* 21645 22359: contig of 715 bp in length
* 22360 22459: gap of 100 bp
* 22460 23153: contig of 694 bp in length
* 23154 23253: gap of 100 bp
* 23254 23941: contig of 688 bp in length
* 23942 24041: gap of 100 bp
* 24042 24741: contig of 700 bp in length

* 24742 24841: gap of 100 bp
* 24842 25541: contig of 700 bp in length
* 25542 25641: gap of 100 bp
* 25642 26348: contig of 707 bp in length
* 26349 26448: gap of 100 bp
* 26449 27069: contig of 621 bp in length
* 27070 27169: gap of 100 bp
* 27170 27873: contig of 704 bp in length
* 27874 27973: gap of 100 bp
* 27974 28665: contig of 692 bp in length
* 28666 28765: gap of 100 bp
* 28766 29463: contig of 698 bp in length
* 29464 29563: gap of 100 bp
* 29564 30231: contig of 668 bp in length
* 30232 30331: gap of 100 bp
* 30332 31031: contig of 700 bp in length
* 31032 31131: gap of 100 bp
* 31132 31820: contig of 689 bp in length
* 31821 31920: gap of 100 bp
* 31921 32597: contig of 677 bp in length
* 32598 32697: gap of 100 bp
* 32698 33328: contig of 631 bp in length
* 33329 33428: gap of 100 bp
* 33429 34132: contig of 704 bp in length
* 34133 34232: gap of 100 bp
* 34233 34936: contig of 704 bp in length
* 34937 35036: gap of 100 bp
* 35037 35743: contig of 707 bp in length
* 35744 35843: gap of 100 bp
* 35844 36557: contig of 714 bp in length
* 36558 36657: gap of 100 bp
* 36658 37362: contig of 705 bp in length
* 37363 37462: gap of 100 bp
* 37463 38172: contig of 710 bp in length
* 38173 38272: gap of 100 bp
* 38273 38978: contig of 706 bp in length
* 38979 39078: gap of 100 bp
* 39079 39782: contig of 704 bp in length
* 39783 39882: gap of 100 bp
* 39883 40577: contig of 695 bp in length
* 40578 40677: gap of 100 bp
* 40678 41372: contig of 695 bp in length
* 41373 41472: gap of 100 bp
* 41473 42185: contig of 713 bp in length
* 42186 42285: gap of 100 bp
* 42286 43021: contig of 736 bp in length
* 43022 43121: gap of 100 bp
* 43122 43824: contig of 703 bp in length
* 43825 43924: gap of 100 bp
* 43925 44631: contig of 707 bp in length
* 44632 44731: gap of 100 bp
* 44732 45447: contig of 716 bp in length
* 45448 45547: gap of 100 bp
* 45548 46258: contig of 711 bp in length
* 46259 46358: gap of 100 bp
* 46359 47062: contig of 704 bp in length
* 47063 47162: gap of 100 bp
* 47163 47862: contig of 700 bp in length
* 47863 47962: gap of 100 bp
* 47963 48665: contig of 703 bp in length
* 48666 48765: gap of 100 bp
* 48766 49445: contig of 680 bp in length
* 49446 49545: gap of 100 bp
* 49546 50252: contig of 707 bp in length
* 50253 50352: gap of 100 bp
* 50353 51034: contig of 682 bp in length
* 51035 51134: gap of 100 bp
* 51135 51840: contig of 706 bp in length
* 51841 51940: gap of 100 bp
* 51941 52641: contig of 701 bp in length
* 52642 52741: gap of 100 bp
* 52742 53422: contig of 681 bp in length
* 53423 53522: gap of 100 bp

	Query Match	3.0%	Score 77:	DB 1:	Length 11286:
	Best Local Similarity	48.0%	Pred. No. 4	1e-09:	
	Matches 320:	Conservative	0:	Mismatches 335:	Indels 12: Gaps 3:
Qy	387	tcattgttggaaagatgcagtgggaagccattggcttggaaatttcagttggaagctgtctg	446		
Db	8788	TCACCATGAACGATGCGCTCGACGCCGGGTGGCGGCGCTATCAACGGCGGTGATGG	8729		

Qy	447	tctctcaactggttggcttgatgcgctgatccatggagatgatcctccagcactctcg	500
Db	8728	GCATATTCCGGCTCGCTCTGGCGCGGATGCCATGCAGGTTCTGGCGGTCCGCTTCACAG	8665
Qy	507	caccacagctgcatctgagtgagagctcccaagctgacagatggcattgctgaacctcg	566
Db	8668	CGGCTCTGATTTCGCGGACCTTTGGCCCTGACAGTGCAGCGAGCGGTGCAGACCGGAAACGG	8609
Qy	567	tggtctttgtaggcatgatctccagctccacagctctctggggaatatctctcagacagctacg	626
Db	8608	CGTTCTTTTTCGGCATGCTGCTGGTGCTGCCGGTTTCGGCGGTTGGCGGACCGCTATG	8549
Qy	627	gcagaaacacaggctaaagatcagcgtctgtggaactgtactataggatcctcttaag	686
Db	8548	GCGCGCGCGGGTGCTGATCGTACCCTGGCTTGGATGCGATGCGTGTTCGCGCGTGTTCGGA	8499
Qy	687	catttgccgcgtgtatagctggtatcctggtgctccggggcctggtgggcttcgggactg	746
Db	8488	TATTTTACCACGATTTCACCATTTCTGCTCGTCTGCTGCTTTCACCGGAGCTCGCGTGG	8429
Qy	747	gagcaggttccccagtcggtg- --acgtgtatgcagagtttcttcccatgaagccagag	803
Db	8428	GCGCACTTTGCGGTGGATTATGCGATGATGGCCGAATTCCTGCTCGCCAAAATCGCG	8369
Qy	804	ctaaatgtatttgcgtgattgaagttattctgggcccattcgggacagttctcgagtlcgctcc	863
Db	8368	GCGCTCGCTGGTGTCTTCTGGAAGGTTTCTGGCGGTTTGGCACACTGATCTGGCACTGG	8309
Qy	864	tggctgtg- ----tctgtatgccccagcttq- --ggctggcggttggctgcacatctct	914
Db	8308	CGGCTTGGGTGCGCAGTCTTGGCGGCGTGGC ;ATGCGCTGGCGGTATATATTTTCGGGTGA	8249
Qy	915	cagctgtcccgctctctctcttctgcgtgctgtgttcttggctgctcgtgaagatgcaaggt	974
Db	8248	CGGCTTTCCGCGCGTCTTGTGCTCGGCGTTCGCTTCTGTCGCCGAATCGCGCTCT	8189
Qy	975	atgatgtcgtgcaggaaaccaggaaaaggcaatcgccaccttaaaaggagatagaactg	1034
Db	8188	ACCTGCTGCGCAGCGGGCGCATCGAAGAGCAAGCGCGCTGCTCAACCGCATGTGTGTGG	8129
Qy	1035	aaaacg 1041	
Db	8128	TGAACGG 8122	
RESULT	14		
AB026645			
LOCUS	81020 bp	DNA	PLN
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MG16, complete sequence.		
ACCESSION	AB026645		07-MAY-1999
VERSION	AB026645.1	GI:4757401	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1 clone:MG16.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids; 1 (sites)		
REFERENCE	Nakamura,K.		
AUTHORS	Structural Analysis of Arabidopsis thaliana Chromosome 3. II		
TITLE	Unpublished (1999)		
JOURNAL	2 (bases 1 to 81020)		
REFERENCE	Nakamura,Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases.		
JOURNAL	Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp. Tel:81-438-52-3935(ex.7443), Fax:81-438-52-3934)		
FEATURES	Location/Qualifiers		

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:46:32 ; Search time 351.36 seconds
(without alignments)
6251.330 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccagcgctccggcttgt.....tgaggactgagngccaagt 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_ll01.*
1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	99.9	2562	21	Human organic cation
2	529.6	20.7	1906	21	Rat organic cation
3	447.4	17.5	480	21	Human 5' EST isola
4	359.8	14.0	401	21	Human secreted exp
5	66.2	2.6	3580	21	Human secreted pro
6	66.2	2.6	4366	22	Human brain cell s
7	63	2.5	2268	22	P. putida oxygenas
8	49	1.9	1356	22	C glutamicum codin
9	49	1.9	309400	22	C glutamicum codin
10	46.6	1.8	1662	19	Human liver cell c
11	46.6	1.8	1888	19	Human liver cell c

12	46.4	1.8	549	21	AAC93874	Cat flea hindgut a
13	46.4	1.8	1590	21	AAC93401	Cat flea HMT synap
c 14	46.4	1.8	1590	21	AAC95402	Cat flea HMT synap
15	46.4	1.8	1875	21	AAC95399	Cat flea HMT synap
c 16	46.4	1.8	1875	21	AAC95400	Cat flea HMT synap
17	44.8	1.7	1519	21	AAC76492	Human OREX ORF2047
18	43	1.7	2191	21	AAH11147	Rat cerebral organ
19	41	1.6	1831	20	AAH26880	DNA encoding a pro
20	41	1.6	1831	21	AAH09889	Human OCTN2 cDNA s
21	40.2	1.6	758	21	AAF15156	Trichoderma reesei
22	40	1.5	2135	20	AAT26879	DNA encoding a pro
23	38.6	1.5	1882	17	AAT08702	Murine organic ani
24	37.6	1.5	1897	22	AAC83979	Streptococcus olea
25	37.6	1.5	50937	21	AAH09469	Human secreted pro
26	37	1.4	470	21	AAC03955	Stealth virus nucl
27	37	1.4	6675	20	AAH84352	Wilson disease gen
c 28	36.6	1.4	744	16	AAH86301	Human OREX ORF1716
29	36.4	1.4	1396	21	AAH76161	Human organic cati
30	36.4	1.4	1490	20	AAH58672	Human saccharide-t
31	36.4	1.4	1560	21	AH248936	Atherosclerosis-as
32	36.4	1.4	1929	22	AAC85075	DNA encoding a pro
33	36.4	1.4	2083	20	AAC26898	Human EST-derived
34	36.4	1.4	2393	22	AAH98580	Human cDNA encodin
35	36.4	1.4	2393	22	AAH22844	Human secreted pro
36	36.4	1.4	2402	21	AH233327	Human organic cati
37	36.4	1.4	2460	20	AAH58671	Human 7TM clone H7
c 38	36.2	1.4	710	22	AAC82672	Novel nucleotide s
c 39	36.2	1.4	839	20	AAH08696	Human olfactory re
c 40	36.2	1.4	949	22	AAH32324	Aspergillus oryzae
41	36	1.4	674	21	AAH12652	Gene encoding a su
42	36	1.4	10732	21	AAH10594	Soybean hexose car
43	35.6	1.4	1960	20	AH232202	Kaposi's sarcoma-a
44	35.6	1.4	3489	21	AAH30290	Nucleotide sequenc
45	35.6	1.4	3489	22	AAH82901	

ALIGNMENTS

RESULT 1
AAZ49686
ID AAZ49686 standard; cDNA; 2562 BP.

XX AAZ49686;

XX 07-APR-2000 (first entry)

XX Human organic cation transporter-like protein (OCTlp) cDNA.

XX Human; organic cation transporter-like protein; OCTlp; transporter;
XX transmembrane; natriuretic; neuroprotective; neuroleptic; anticonvulsant;
XX antiParkinsonian; antidepressant; cellular process; cell proliferation;
XX screen; treatment; prevention; diagnosis; neurodegenerative disorder;
XX Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
XX CNS disorder; central nervous system; schizophrenia; depression;
XX behavioural; sleep disorder; eating disorder; Alzheimer's; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 185..1831

XX FT /*tag= a

XX FT /product= "OCTlp"

XX PN WO200000633-A1.

XX PD 06-JAN-2000.

XX PF 29-JUN-1999; 99WO-US14880.

XX PR 30-JUN-1998; 98US-0107932.

XX PA (MILL-) MILLENNIUM PHARM INC.

|||||
Db 1681 cctgctgcctgctcctctgttttggccatgagaccacaaagcgagagtcagga 1740
QY 1741 gtccagccacgggagtggtggccaggaagatggtcgccgaggaatgacagtgaggtgt 1800
Db 1741 gtccagccacgggagtggtggccaggaagatggtcgccgaggaatgacagtgaggtgt 1800
QY 1801 taccaggtcgaaactctgctctcaggaatagtgacagtggtgggagactgagctgtctttg 1860
Db 1801 taccaggtcgaaactctgctctcaggaatagtgacagtggtgggagactgagctgtctttg 1860
QY 1861 aggctgcagagcttggggggtggcagggcccaactggggcactgattgtcactgccagc 1920
Db 1861 aggctgcagagcttggggggtggcagggcccaactggggcactgattgtcactgccagc 1920
QY 1921 atcaagaactcaccacaaagatagacctggaccacaggggttttggctctgactcagtt 1980
Db 1921 atcaagaactcaccacaaagatagacctggaccacaggggttttggctctgactcagtt 1980
QY 1981 tgcctatcttcattgaggtccaccagggatggggagatgtttgctctaggggttctct 2040
Db 1981 tgcctatcttcattgaggtccaccagggatggggagatgtttgctctaggggttctct 2040
QY 2041 gatatgtggtgaaagcttggcttaaccctggtgatctacatggaagactaccatatt 2100
Db 2041 gatatgtggtgaaagcttggcttaaccctggtgatctacatggaagactaccatatt 2100
QY 2101 taggaggtctgtaatgccagcaaccacaaatcagacaccacacagagtcacccggccaaac 2160
Db 2101 taggaggtctgtaatgccagcaaccacaaatcagacaccacacagagtcacccggccaaac 2160
QY 2161 cctcagtgaaacacacaaaatatctctctgtagataccgtccaggtccagccatgtgac 2220
Db 2161 cctcagtgaaacacacaaaatatctctctgtagataccgtccaggtccagccatgtgac 2220
QY 2221 acctgtctccaccacagacctgttcagtaggtttctccacacccacagccccagc 2280
Db 2221 acctgtctccaccacagacctgttcagtaggtttctccacacccacagccccagc 2280
QY 2281 ttctctttgaaatgcagggcgatagtggtggtctgtagcagctatttctgagcagg 2340
Db 2281 ttctctttgaaatgcagggcgatagtggtggtctgtagcagctatttctgagcagg 2340
QY 2341 gcccccgtttgctctcctagagcctgaccagtgattctctggcagatgacatgggtg 2400
Db 2341 gcccccgtttgctctcctagagcctgaccagtgattctctggcagatgacatgggtg 2400
QY 2401 cattcaactggagccacatgccccacccagccccctnttggagttgcccgtttgtggca 2460
Db 2401 cattcaactggagccacatgccccacccagccccctnttggagttgcccgtttgtggca 2460
QY 2461 ccaagagatccagatgtctcctgggacacagctgggtctctgaccaggtgacacacctcaaa 2520
Db 2461 ccaagagatccagatgtctcctgggacacagctgggtctctgaccaggtgacacacctcaaa 2520
QY 2521 acgcccgttaccctctggggaactgagactgagngcgaagt 2562
Db 2521 acgcccgttaccctctggggaactgagactgagngcgaagt 2562

RESULT 2
AAZ49689
ID AAZ49689 standard; cDNA; 1906 BP.
XX
AC AAZ49689;
XX
DT 07-APR-2000 (first entry)
XX
DE Rat organic cation transporter-like protein (OCTlp) partial cDNA.
XX
KW Human; organic cation transporter-like protein; OCTlp; transporter;
KW transmembrane; norepinephrine; neuroprotective; neuroleptic; anticonvulsant;
KW antiparkinsonian; antidepressant; cellular process; cell proliferation;

KW screen; treatment; prevention; diagnosis; neurodegenerative disorder;
KW Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
KW CNS disorder; central nervous system; schizophrenia; depression;
KW behavioural; sleep disorder; eating disorder; rat; ss.
OS Rattus sp.
XX WO200000633-A1.
XX 06-JAN-2000.
XX 29-JUN-1999; 99WO-US14880.
XX 30-JUN-1998; 98US-0107932.
XX (MILL-) MILLENNIUM PHARM INC.
XX Goodearl AJ, Glucksmann MA;
XX WPI; 2000-137069/12.
XX
XX New nucleic acid encoding human organic cation transporter-like
XX protein, used for prevention, treatment and diagnosis of e.g.
XX neurological, behavioural or sleep disorders
XX Disclosure; Fig 4; 100pp; English.
XX
XX The present sequence is a partial cDNA encoding rat OCTlp (organic
XX cation transporter-like protein) and derived from rat frontal cortex
XX foetal brain cDNA library. Human OCTlp is a member of the superfamily of
XX sugar and other transporter molecules that have 12 transmembrane
XX domains. It is highly expressed in brain tissue and has norepinephrine,
XX neuroprotective, norepinephrine, anticonvulsant, antiparkinsonian,
XX antidepressant activities. The OCT-like protein is used to
XX regulate a variety of cellular processes e.g. cell proliferation,
XX differentiation and survival, screen OCTlp modulators and detect mutation
XX in OCTlp gene. OCTlp modulators can be used to treat or prevent chronic
XX neurodegenerative disorders (e.g. Alzheimer's, Parkinson's,
XX Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic,
XX depression), behavioural, sleep and eating disorders.
XX
XX Sequence 1906 BP; 379 A; 547 C; 556 G; 424 T; 0 other;

Query Match 20.7%; Score 529.6; DB 21; Length 1906;
Best Local Similarity 73.9%; Pred No. 2.7e-138;
Matches 800; Conservative 0; Mismatches 254; Indels 28; Gaps 9;
QY 1286 ctgagtgaggaggtatgacagtgactgtgtgacacccctctctgagttccaggtgtc 1345
Db 15 ccgagcaagaggtatgacagtgactgtgtgacacccctctctgagttccaggtgtc 74
QY 1346 ctgtgactctgtgattatgacagctggggcgacagagacacacagctgtgtctt 1405
Db 75 ctgtgactctgtggttcacagcgcctggggcgacagagacacacagctgtgtctt 134
QY 1406 gtcaatctctctctgacgctcctgtgtttatctgtgttggagaatgtgtcact 1465
Db 135 gtcaatctctctctgacgctcctgtgtttatctgtgttggagaatgtgtcact 194
QY 1466 ctgttactctctgacagagcgtttatttctgaggtttcaagcgcatattttac 1525
Db 195 ctcttactgtcttgcagagcgtttatttctgaggtttcaagcgcatattttac 254
QY 1526 acactgaggtctaccccgacggcgccctgctgctggcgacacacacacacacac 1585
Db 255 acgctgaggtgtatccacggcgacagggcgctggcgacacacacacacacac 314
QY 1586 gcaagagtggtgtctatcactcctgtctcactgcgaggtgatgtggaactcctgtg 1645
Db 315 gcgagagtggtgtctatcactcctgtctcactgcgaggtgatgtggaactcctgtg 374

```
QY 1646 tacctgaactctgacgtttacagtgcgtcgtcctcctggtcgtcctcctcctgcttt 1705
Db 375 tacctgaactctgacgtttacagtgcgtcctcctcctcctcctcctcctcctcctt 434
QY 1706 ttgccattgagacaaagcggagactgcaggagtcacagccacggagtgggccacag 1765
Db 435 ctgccatcgacacaaagccgacactgcaggagtcacagccacggagtgggccacag 494
QY 1766 gagatggtcggccaggaaatgcacggtgcagggtgtaccaggtcgaactctgctctcag 1825
Db 495 gagatggttggccagggaacacagcagcagcgtcccccagtcgaactctgctctcag 554
QY 1826 gaatagtcacgacagggactgagctgtgtcttgcagcagcagcgttggggcctgcgc 1885
Db 555 gagtgtgacctcctggag-ttagctgtgtcttgcagc--cgagcctagaagactgctgc 611
QY 1886 agcccccaactgggacgattgtcactgcgcacatacaagaac-tcacccaagagtatg 1944
Db 612 agagccacgctggccactaacggtcactgcgcacatacaagaacttccccagtggtgcg 671
QY 1945 acctgacacaaaggtttgtgtcttgcactcagttt-gctcattctcattgagg----- 1998
Db 672 aagtgaacacaggggtttgtgtctcactgtgttggcctatgttccatcaggggttgc 731
QY 1999 --tccaccacgaggtggagatgtttgtctaggggttctctgtatagtgtggaag 2056
Db 732 ccgcccaggaagatgggctgattcactcctcaggggttctcgtggtggggaagagg 791
QY 2057 ctttgtcataactgtgatatcagtggaagactaccataattaggagggtctgtgtaa 2116
Db 792 gtgtgacgtgcgc-gtggatctgcattggggaagctgctagtgtggagggtccacagg 850
QY 2117 tgcacgaacacatcagacaccacaga--gtcacccggccaaacccctcagtgaaacac 2174
Db 851 cgtcagggccagctgagcagatgtcacgtgtgtaccagtcatacccttgagagccac 910
QY 2175 -caaaatctctctatagataccgtccaggtcagggccatctgacacactgctgtccac 2233
Db 911 tgtccaaagatcccatagatacagctcagccagaccctgtgacaccccat---- 966
QY 2234 ccaccgacgttcagtaggtttctccacacccacagcccgaggtcttctcttgaa 2293
Db 967 -----ttgtccagtaagttcttccgcacccctgcccaggtgcttttggaatta 1018
QY 2294 attgacgcatctagtggtctgagcagctatttctcgtggcaggggccccccggttg 2353
Db 1019 agacaagtaattagtgccgactagacagcttttctgagcctgagacaccccccttc 1078
QY 2354 cc 2355
Db 1079 cc 1080

RESULT 3
AAZ42410
ID AAZ42410 standard; cDNA; 480 BP.
XX
AC AAZ42410;
XX
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:169.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW gene therapy; chromosome mapping; upstream regulatory sequence;
KW forensic; location; development; protein synthesis; stability;
KW regulation; identification; ss.
XX
OS Homo sapiens.
XX
XX WO9553051-A2.
XX
PD 21-OCT-1999.
```

```
XX
PF 09-APR-1999; 99WO-IB00712.
XX
PR 09-APR-1998; 98US-0057719.
PR 28-APR-1998; 98US-0069047.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-038446/03.
DR P-PSDB; AAY64796.
XX
PT Novel secreted protein 5' expressed sequence tag sequences used in
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS Claim 1; Page 265-266; 837pp; English.
XX
XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
CC sequences, corresponding to human secreted proteins. AAY64651 to
CC AAY65438 represent the EST-related proteins corresponding to AAZ42265 to
CC AAZ43052. The 5' ESTs can be used for producing secreted human gene
CC products. They can be used to identify and isolate 5' untranslated
CC regions (UTRs) and upstream regulatory regions which control the
CC location, development stage, rate, and quantity of protein synthesis, as
CC well as stability of mRNA. The ESTs are also useful as probes for
CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC also be used in forensic procedures to identify individuals, or in
CC diagnostic procedures to identify individuals having genetic diseases
CC resulting from abnormal gene expression. The products may also be used in
CC gene therapy protocols. The nucleic acids encoding signal peptides can be
CC used for directing extracellular secretion of a polypeptide or the
CC insertion of a polypeptide into a membrane, or importing a polypeptide
CC into a cell. The proteins encoded by the EST sequences may be useful in
CC treating a variety of human conditions. Secreted proteins have
CC therapeutic value, and the identification of new secreted proteins is
CC valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 480 BP; 102 A; 115 C; 154 G; 100 T; 9 other;

Query Match 17.5%; Score 447.4; DB 21; Length 480;
Best Local Similarity 97.7%; Pred. No. 1.6e-115;
Matches 464; Conservative 9; Mismatches 0; Indels 2; Gaps 2;

QY 19 gtctcgtgggtcgtcccccagcggctccctcccccagcagctgctcctcagggaggaag 78
Db 1 gtctcgtgggtcgtcccccagcggctccctcccccagcagctgctcctcagggaggaag 60
QY 79 cggcgcgggtcgtcctcagcttcccggtgctgataaaacccgagggctcgtcctccaccact 138
Db 61 cggcgyrrgtgctcagcttcccggtgctgataaaacccgagggctcgtcctccaccact 120
QY 139 accatgaaggccatgagaaggcctcctcctcctcctcctcctcctcctcctcctcctcct 198
Db 121 accatgaaggccatgagaaggcctcctcctcctcctcctcctcctcctcctcctcctcct 179
QY 199 attcagctaaaggcagctcccggttgcgtgataaaattccgcagcagcagagtgcaaggtc 258
Db 180 attcagctaaaggcagctcccggttgcgtgataaaattccgcagcagcagagtgcaaggtc 239
QY 259 agaggcagcacgcgcttcagagagcatgaagtccagattgaagggttccacgtggtgcct 318
Db 240 agaggcagcacgcgcttcagagagcatgaagtccagattgaagggttccacgtggtgcct 299
QY 319 agaggcgtggaagctggatgatgggcagcagctgtgcccaaggagtttgcacaatccaccga 378
Db 300 agaggcgtggaagctggatgatgggcagcagctgtgcccaaggagtttgcacaatccaccga 359
QY 379 tgatacttccatggtggaagatcagtggaagcattgcttgggaaatttccagtggaa 438
Db 360 tgatacttccatggtggaagatcagtggaagcattgcttgggaaatttccagtggaa 419
```

QY 439 gctgtctgtctcactggtgcttgatggctgagtgccatgagatgatgac 493
 |||||
 Db 420 gctgtctgtctcactggtgcttgatggctgagtgccatgagatgatgac 473
 |||||

RESULT 4
 AAA45312
 ID AAA45312 standard; cDNA; 401 BP.
 XX
 AC AAA45312;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:1887.
 XX
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
 KW antitumor; osteopathic; neuroprotective; neurotropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX
 PR 15-OCT-1998; 98US-0104436.
 XX
 PA (GEM) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 XX
 DR WPI: 2000-317938/27.
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (sESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 PS Claim 1: Page 679; 803pp; English.
 XX
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The sESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytosolic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;
 CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The sESTs can be used for gene
 CC therapy and in vaccines. The sESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation

CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.

XX
 SQ Sequence 401 BP; 69 A; 119 C; 104 G; 109 T; 0 other;

Query Match 14.0%; Score 359.8; DB 21; Length 401;
 Best Local Similarity 98.1%; Pred. No. 5.5e-91;
 Matches 364; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1386 agaccatggccctgtgcttctcattctctctcctcagcctcctgctgtttatctgtg 1445
 |||||

Db 31 agaccatggccctgtgcttctcattctctcctcagcctcctgctgtttatctgtg 90
 |||||

QY 1446 ttggaagaaatgtgctcactctgttactcttcattgcaagagcggtttatctggaggt 1505
 |||||

Db 91 ttggaagaaatgtgctcactctgttactcttcattgcaagagcggtttatctggaggt 150
 |||||

QY 1506 ttcaagcgccatattgtttacacacctgaggtctacccacgcaacgagcgccctcgcc 1565
 |||||

Db 151 ttcaagcgccatattgtttacacacctgaggtctacccacgcaacgagcgccctcgcc 210
 |||||

QY 1566 tggcaccctgcagcgccatggcgaagtggtgctctcactcctcctcctcctcctcag 1625
 |||||

Db 211 tggcaccctgcagcgccatggcgaagtggtgctctcactcctcctcctcctcctcag 270
 |||||

QY 1626 tgatgctggaatcctctgtgtacctgacctgagcttgagcttgagctgctgctcctg 1685
 |||||

Db 271 tgatgctggaatcctctgtgtacctgacctgagcttgagcttgagctgctgctcctg 330
 |||||

QY 1686 ctgcccctggcctcctgcttttggccattgagacaaaggcgaggagactgcaggagtcca 1745
 |||||

Db 331 ctgcccctggcctcctgcttttggccattgagacaaaggcgaggactgcaggagtcca 390
 |||||

QY 1746 gccaccgggag 1756
 |||||

Db 391 agccactcgag 401

RESULT 5

AAA16697

ID AAA16697 standard; cDNA; 3580 BP.

XX

AC AAA16697;

XX

DT 16-JUN-2000 (first entry)

XX

DE Human secreted protein clone as180_1 nucleotide sequence SEQ ID NO:159.

XX

KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy; ss.

XX Homo sapiens.

OS

XX WO200009552-A1.

PN

XX 24-FEB-2000.

PD

XX 13-AUG-1999; 99WO-US18298.

PF

XX 14-AUG-1998; 98US-0096622.

PR 17-AUG-1998; 98US-0096815.

PR 04-SEP-1998; 98US-0099229.

SQ Sequence 4366 BP; 916 A; 1187 C; 1195 G; 1068 T; 0 other;									
Query Match 2.6%; Score 66.2; DB 22; Length 4366;									
Best Local Similarity 47.3%; Pred. No. 5.1e-08;									
Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;									
Qy	415	tggcttggaaaatttcagtggaagctgtctgttctcaactggttgcttggatggctga	474						
Db	911	tggccagcgctgtccagtgacactgtatttctgttctggtctgctgctgctgctga	970						
Qy	475	tggcctggagatgatctctcagcatctctggcaccacagctgcctgcagatggagct	534						
Db	971	cgggtggaggtcttctgggtggtctctgtctcccgagcgtgagaaadacatgtgctt	1030						
Qy	535	cccaagctggcaggtggcattgctgacctcgtgtgttctttagcctgatgtccagctc	594						
Db	1031	gtccgaactcccaaaaaggcatgctaggcctcatcgtctcaactgggcagatggtgggagc	1090						
Qy	595	cacgctctggggaaatatctcagaccagtagcggcaggagaaacagggtgaaagatcacgcgt	654						
Db	1091	cttctctcggggaggtctgctgaccgctggtgcggaggcagtgctgtctcatctcgt	1150						
Qy	655	gctgtggaactctgtactatgacatccttagtgcaatttcgcccgtgtatagctggaacct	714						
Db	1151	ctcagtcacacagctctctgcctctctctcattcttcttgcagggttcacggcaactttcct	1210						
Qy	715	ggctgctccggggctgggtcttcggatcgaggg---agttccccagtcggtgacgct	771						
Db	1211	ctcttcgcgcctactcttctgggttggaatggaggtccatcccatcttctctcta	1270						
Qy	772	gtatcccgagttcctcccatgaaagccagagctaaatgtattttgtgtatgaggtatt	831						
Db	1271	tttctccgagttctggccccaggagaaacaggagcatttgagctggctctcatgtt	1330						
Qy	832	ctgggccatcgagacagatttcagggtcgtcctggtgtgtctgtgtgagtcgccagctggg	891						
Db	1331	tggagatgttggggcgtgtacgagctgtatggcctggggccatcccccactatgg	1390						
Qy	892	ctggcgttggctgt 906							
Db	1391	gtggagtttccagat 1405							
RESULT 7									
ID	AAF26353	AAF26353 standard; DNA; 2268 BP.							
XX	AC	AAF26353;							
XX	DT	02-MAY-2001 (first entry)							
XX	DE	P. putida oxygenase encoding DNA ORF04677.							
XX	KW	Oxygenase; transgenic plant; detection; amino acid production;							
XX	KW	vitamin production; steroid production; dialcohol production;							
XX	KW	oxidized aromatic compound production; dialdehyde production;							
XX	KW	optically active epoxide production; dicarboxylic acid production;							
XX	KW	bioconversion; ORF04677; ds.							
OS		Pseudomonas putida.							
XX	PN	WO200107629-A2.							
XX	PD	01-FEB-2001.							
XX	PF	27-JUL-2000; 2000WO-EP07244.							
XX	PR	27-JUL-1999; 99DE-1035087.							
XX	PA	(TIGR-) TIGR INST GENOMIC RES.							
XX	PA	(QUT-) QUTAGEN GMBH.							
XX	PA	(GBF) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.							

(DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
(MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

Fraser CM, Venter C, Tuemmler B, Hoheisel J, Duesterhoeft A;
Hilbert H, Timmis KN, Moore E, Straetz M, Heim S;
WPI; 2001-168556/17.

New DNA encoding a bacterial oxygenase, useful e.g. for production of
amino acids or vitamins, also derived transgenic plants -
Claim 3a; Page 35-36; 47pp: German.

This invention describes novel DNA sequences (I) that express products
having the biological function of oxygenases. The invention also
describes (I) recombinant expression vectors containing (I); (2)
prokaryotic and eukaryotic cells transformed or transfected with (I) or
the vector of (I); (3) production of oxygenases by culturing cells of
(2); (4) (partial) expression products (II) of (I), and synthetic
proteins or peptides with the same sequences; (5) mono- or poly-clonal
antibodies (Ab) specific for (II); (6) hybridoma cells that produce
monoclonal Ab; and (7) transgenic plants that contain cells of (2). (I),
and their fragments, are useful for the following: (i) expression of
recombinant oxygenases; (ii) useful as probes and primers for detection,
isolation and amplification of full-length cDNA sequences; and (iii) used
to produce transgenic plants. Expression products are useful for
production of amino acids, vitamins, steroids, oxidized aromatic
compounds, optically active epoxides, dialcohols, dialdehydes and
dicarboxylic acids. The transformed cells that express them are used for
bioconversion of aromatic and aliphatic compounds. The oxygenases have
very weak substrate specificity so can be used for highly stereo- or
regio-specific oxidations of a wide range of substrates, reactions that
are difficult or impossible to do with conventional oxidizing agents.

SQ Sequence 2268 BP; 371 A; 713 C; 732 G; 450 T; 2 other;

Query Match 2.5%; Score 63; DB 22; Length 2268;
Best Local Similarity 45.9%; Pred. No. 2.9e-07;
Matches 253; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

Qy	417	gctttggaaaaatttcagtggaagctgtctgtctcactggttgcttggttgcttgatg	476						
Db	41	gctttaccctgttctcactggatggttcagtggtgctgctgctgctgctgttttcgacg	100						
Qy	477	ccatggagatgatgactcctcagcatctcctggcaccacagctgcatgctgagtgagctcc	536						
Db	101	gctatgacctgttcatctacgtgtggtgactgctgctgctgctgctgctgctgctgctg	160						
Qy	537	caagctggcagtggtgctgacctggtgactggtgttctttaggagctgctgctgctgctg	596						
Db	161	ccccgttgacggtggtgctgacctggcagctgctgctgctgctgctgctgctgctgctg	220						
Qy	597	cgctctggggaaaatactcagaccagtagcgcagagaaacagggctgaagatcacgctgc	656						
Db	221	tggcctcggcagctggtccagcctgctgctgctgctgctgctgctgctgctgctgctg	280						
Qy	657	tgtggaactctgactatggtcattcttagtgcatttgcgcctggtatagctggatccctgg	716						
Db	281	cttgttctcggggcagcaccatctcctcaatgctgttgccagcaacccgagagtttgcca	340						
Qy	717	tgctccgggggctggtggtggtcggagctgagagaggtt---ccccagctcggtgagctgt	773						
Db	341	tctaccgcttcatcgcggctggtggtggtggtggtggtggtggtggtggtggtggtggtg	400						
Qy	774	atgcgaggttctctcccatgaaagccagagctaaaatgtatttgcgtgattgaggtattct	833						
Db	401	tgaacgatacgcacccaagcctgcgcagcagcactggtggtgcgcatctgttcagtgtgt	460						
Qy	834	gggccatcgggacagtgttctcaggtgctcctggtggtggtggtggtggtggtggtggtg	893						
Db	461	attcgtggggcgcagctgctgcggcagggtgtcggcatcttcattcgtcgcgctgtttggt	520						

CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;
SQ

Query Match 1.9%; Score 49; DB 22; Length 309400;
Best Local Similarity 45.5%; Pred. No. 0.032;
Matches 215; Conservative 0; Mismatches 255; Indels 3; Gaps 1;
QY 473 gatgccatgagatgatctctcagcattcctgagccacagctgcatgctgagtgaggy 532
DB 227907 gacggctacacacattgtgtacgacccgtccaatcgccctgctgaaggatggaac 227966
QY 533 ctcccaagcggcaggtggcattgctgacccgtggtgctctttgagcattgattccagc 592
DB 227967 ttaagctctcaacgctcgggcaccattgctccacgcgttctttggcattgctgacgc 228026
QY 593 tccacgctctgggaaatatctcagaccagtcagcaggaggaacacaggcgtgaagatcagc 652
DB 228027 gctgtgtcattgtctgactgtcagaccgctggtggccgaaagcagcgtgattggtacc 228086
QY 653 gtgctgtggactgtactatgcatctcattagtcatttcgcccgtgtatagctggtac 712
DB 228087 gtgctgtattctctctctcaccattctgtgtcatttctcctcaaacccattgggtctc 228146
QY 713 ctggtgctccgggctggtggtggtcctggatcgagagagatttcccagtcggtga---cg 769
DB 228147 ggcgttttcgcttaccagcagcctggtctggtggcgttggcttcagttaatgctg 228206
QY 770 ctgtatccgaggttctctccatgaaagccagagactaaatgtatttctgattgaggtta 829
DB 228207 atgacctctgatttggcttccacgttaagaccattcagcgtgggcaacggtgatgctcg 228266
QY 830 ttcctggccatcgagcagtggttcgaggtgctgctgctgctgctgctgctgctgctg 889
DB 228267 ggcgttcccatcggtgttctatcgccgctgctgctgctgctgctgctgctgctgct 228326
QY 890 ggcctggcgtggtgctcctcctcagctgctccgctcctcctcctcctcctcctcct 942
DB 228327 gaggagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 228379

RESULT 10
AAV49559
ID AAV49559 standard; cDNA to mRNA; 1662 BP.
XX
AC AAV49559;
XX
XX
DT 21-Oct-1998 (first entry)
XX
DE Human liver cell clone HP01293 cDNA #2.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX

PR 13-NOV-1996; 96JP-0301429.
XX
XX (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX
XX WPI; 1998-297932/26.
DR P-PSDB; AAW64538.
XX
XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX
XX Claim 3; Page 120-121; 205pp; English.
PS
XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
CC a transmembrane domain. These proteins can be used for, e.g. research
CC and nutrition, and may have cytokine and cell
CC proliferation/differentiation, immune stimulating/suppressing,
CC haematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 1662 BP; 312 A; 486 C; 466 G; 398 T; 0 other;

Query Match 1.8%; Score 46.6; DB 19; Length 1662;
Best Local Similarity 47.8%; Pred. No. 0.0097;
Matches 232; Conservative 0; Mismatches 244; Indels 9; Gaps 3;
QY 1267 cagcctggcctgcagtcactgagtgagggagattacatgactgtgtggaccacct 1326
DB 1092 catctgcacatggggcaccagcgggaacctcactgatttcttactccgctct 1151
QY 1327 ctctgatttctcaggtgtcctgtgacctgtggtgatttattgacccgtggggcgcaagaa 1386
DB 1152 ggtcgaatccggggcctctcatagccctcaccattgacgcgtggggcgcatcta 1211
QY 1387 gaccatggcctgtgcttctgctctctct---cctctgcagcctcctgctgtttatctg 1443
DB 1212 cccatggcgtgctcaaatgttgccggggcagcctgctgctgattttatctc 1271
QY 1444 tgttggagaataatgctcactctgttactcttca----ttgcaagagcgtttattctgg 1500
DB 1272 acctgacctgcactggttaaacatcataatcatgtgtgtggcgaatgggaatcaccat 1331
QY 1501 agcctttcaagcgcacatattttacacacctgaggtctaccacacggcgaacgcggacct 1560
DB 1332 tgcatacaaatgatcgcctggtggaatgctgagctgtaccacacattcgcaagaaacct 1391
QY 1561 cggcctgggcacctgcagcggcagtggaagagtggtgtctctcatcactcctgttcacgc 1620
DB 1392 cggagtgatggtgtgttctcctcctgtgtgacataggcgggataatcaccccttcacagt 1451
QY 1621 ccaggtgatgctggaatcctctgtgtacacctgactctggaatttaccagtgctgctcct 1680
DB 1452 ctccaggtgagggaggtcctggaagccttgccctcatttctgttgcgggtgtggtggcct 1511
QY 1681 cctggtgctcctggcctcctcctcttcttggccattgagaccacacggcgagactgcagga 1740
DB 1512 gcttgcgcgggaagtgcgctacttcttcca--gagaccaagggggtcgtcttggccaga 1568
QY 1741 gtcca 1745
DB 1569 gacca 1573

RESULT 11
AAV49558
ID AAV49558 standard; cDNA to mRNA; 1888 BP.
XX
AC AAV49558;
XX

SQ Sequence 549 BP; 162 A; 83 C; 122 G; 182 T; 0 other;

CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic acids which are expressed in hindgut and Malpighian tubule (HMT) tissue

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:45:17 ; Search time 137.14 Seconds
(without alignments)
4230.980 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccacgcgtccggcttctgt.....tgaggactgagngccaagtq 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.2	2.1	1885	3	US-08-501-572-5
2	53.2	2.1	1885	3	US-09-040-444-5
3	45.2	1.8	7218	1	US-08-232-463-14
4	45	1.8	1896	3	US-08-501-572-6
5	45	1.8	1896	3	US-09-040-444-6
6	42	1.6	7218	1	US-08-232-463-14
7	38.6	1.5	1882	3	US-08-501-572-4
8	38.6	1.5	1882	3	US-09-040-444-4
9	37.6	1.5	50937	4	US-09-428-517-1
10	36.4	1.4	1931	2	US-09-130-114-2
11	36.4	1.4	1490	4	US-08-964-127-3
12	36.4	1.4	2460	4	US-08-964-127-1
13	35.6	1.4	3489	2	US-08-728-323A-1
14	35.6	1.4	32207	2	US-08-770-379-20
15	35.6	1.4	32207	4	US-08-757-669A-20
16	35.2	1.4	2496	1	US-08-073-384C-2
17	35.2	1.4	2496	1	US-08-254-359A-2
18	35.2	1.4	2496	1	US-08-483-043-2
19	35.2	1.4	2496	1	US-08-481-238-2
20	35.2	1.4	2496	2	US-08-471-066B-2
21	35.2	1.4	2496	2	US-08-484-956-2
22	35.2	1.4	2496	2	US-08-757-653-2
23	35.2	1.4	2496	2	US-08-599-491-2
24	35.2	1.4	2496	2	US-08-756-386-2
25	35.2	1.4	2496	2	US-08-823-516-2
26	35.2	1.4	2496	3	US-08-682-853A-2
27	35.2	1.4	2496	3	US-08-759-038-2

28	35.2	1.4	2496	3	US-08-758-314-2	Sequence 2, Appli
29	35	1.4	2082	4	US-09-440-325A-2	Sequence 2, Appli
30	34.8	1.4	2102	2	US-08-647-397-1	Sequence 1, Appli
31	34.4	1.3	289	4	US-09-007-005-17	Sequence 17, Appli
32	34.4	1.3	289	4	US-09-244-796-17	Sequence 17, Appli
33	34.2	1.3	2484	4	US-09-276-531-46	Sequence 46, Appli
34	34	1.3	2515	1	US-08-061-465-3	Sequence 3, Appli
35	34	1.3	6828	1	US-08-061-465-1	Sequence 1, Appli
36	33.6	1.3	803	1	US-07-928-611-12	Sequence 12, Appli
37	33.6	1.3	803	2	US-08-487-811A-12	Sequence 12, Appli
38	33.6	1.3	803	4	US-09-060-694-12	Sequence 12, Appli
39	33.6	1.3	803	5	PCT-US93-07370-12	Sequence 12, Appli
40	33.6	1.3	1610	1	US-08-056-051-5	Sequence 5, Appli
41	33.6	1.3	1610	1	US-07-928-611-21	Sequence 21, Appli
42	33.6	1.3	1610	2	US-08-487-811A-21	Sequence 21, Appli
43	33.6	1.3	1610	4	US-09-060-694-21	Sequence 21, Appli
44	33.6	1.3	1610	5	PCT-US93-07370-21	Sequence 21, Appli
45	33.6	1.3	2019	4	US-09-063-950-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-08-501-572-5

: Sequence 5, Application US/08501572

: Patent No. 6063623

: GENERAL INFORMATION:

: APPLICANT: Koepsell, Hermann

: APPLICANT: Grundeman, Dirk

: APPLICANT: Gorboulev, Valentin

: TITLE OF INVENTION: Transport protein Which Effects The

: TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,

: TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.

: NUMBER OF SEQUENCES: 6

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner

: STREET: 1300 I Street, N.W., Suite 700

: CITY: Washington

: STATE: D.C.

: COUNTRY: USA

: ZIP: 20005-3315

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent In Release #1.0, Version #1.30

: CURRENT APPLICATION DATA: US/08/501.572

: FILING DATE:

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Toohey, Kimberlin M

: REGISTRATION NUMBER: 35,391

: REFERENCE/DOCKET NUMBER: 02481.1453-00000

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (202)408-4000

: TELEFAX: (202)408-4400

: INFORMATION FOR SEQ ID NO: 5:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1885 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

US-08-501-572-5

Query Match 2.1%; Score 53.2; DB 3; Length 1885;

Best Local Similarity 47.3%; Pred. No. 1e-05;

Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

Oy 1267 cagcctgcgcgagctgagtgaggagattacatggacttgctgtggaccacct 1326

```

Db 1164 CATCTGACATGGCGGCACACAGGGAACCTCTACCTGGATTTCCTTTACTCCGCTCT 1223
Qy 1327 clctgagttccagtgctctgtgactctgtgattattgaccgctgggagcgcaagaa 1386
Db 1224 GGTGGAATCCCGGGGCTTCATAGCCCTCATACCATTTGACCGCGTGGCGCATCTA 1283
Qy 1387 gaccatggccctgtgcttctgtatctctctctctctctctctctctctctctctct 1446
Db 1284 CCCATGGCGGTGCAAAATTTGTTGGCGGGGACAGCTGCCCTCGTCATTTTATCTCACC 1343
Qy 1447 tgaagaaatgctcactctgtactctca---ttcaagagcgtttatttctggagg 1503
Db 1344 TGACCTGCACTGGTTAAACATCATATCATGTGTGTCGGAATGGAATCACCATTGC 1403
Qy 1504 cttcaagcggcagatgtttacacacctgaggtctacccccagcgcaacgcgccctcgg 1563
Db 1404 AATACAAATGATCTGCCCTGGTGAATGCTGAGCTGTACCCACATTCGTCAGGAACCTCAG 1463
Qy 1564 cctgggacccctcagcgagtgcaagagtggtgctctctctctctctctctctctctc 1623
Db 1464 AGTGATGGGTGTTCTCCCTGTTGACATAGGTGGGATAATCACCCCTTCATATAGTCTT 1523
Qy 1624 ggtgatgctggaatcctctgtgactgactctgagctgtgactgagctgctcct 1683
Db 1524 CAGGCTGAGGAGGCTGCGAAGCCTTGCCCTCATTTTGTGGGTGTTGGGCTGCT 1583
Qy 1684 ggtctgctgctcctctctctctctctctctctctctctctctctctctctctct 1743
Db 1584 TGCCCGGGAGTGAGCGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
Qy 1744 ca 1745
Db 1641 CA 1642

RESULT 2
US-09-040-444-5
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5
```

```

Query Match 2.1%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 1e-05;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
```

```

Qy 1267 cagcctggcctgcgagctacgtgagtgagagattacatggaactgctgtagcaccct 1326
Db 1164 CAICCTGCACATGGCGGCACACAGGGAACCTCTACCTGGATTTCCTTTACTCCGCTCT 1223
Qy 1327 clctgagttccagtgctctgtgactctgaggtatttaccgctgggagcgcaagaa 1386
Db 1224 GGTGGAATCCCGGGGCTTCATAGCCCTCATACCATTTGACCGCGTGGCGCATCTA 1283
Qy 1387 gaccatggccctgtgcttctgtatctctctctctctctctctctctctctctctct 1446
Db 1284 CCCATGGCGGTGCAAAATTTGTTGGCGGGGACAGCTGCCCTCGTCATTTTATCTCACC 1343
Qy 1447 tgaagaaatgctcactctgtactctca---ttcaagagcgtttatttctggagg 1503
Db 1344 TGACCTGCACTGGTTAAACATCATATCATGTGTGTCGGAATGGAATCACCATTGC 1403
Qy 1504 cttcaagcggcagatgtttacacacctgaggtctacccccagcgcaacgcgccctcgg 1563
Db 1404 AATACAAATGATCTGCCCTGGTGAATGCTGAGCTGTACCCACATTCGTCAGGAACCTCAG 1463
Qy 1564 cctgggacccctcagcgagtgcaagagtggtgctctctctctctctctctctctctc 1623
Db 1464 AGTGATGGGTGTTCTCCCTGTTGACATAGGTGGGATAATCACCCCTTCATATAGTCTT 1523
Qy 1624 ggtgatgctggaatcctctgtgactgactctgagctgtgactgagctgctcct 1683
Db 1524 CAGGCTGAGGAGGCTGCGAAGCCTTGCCCTCATTTTGTGGGTGTTGGGCTGCT 1583
Qy 1684 ggtctgctgctcctctctctctctctctctctctctctctctctctctctctct 1743
Db 1584 TGCCCGGGAGTGAGCGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
Qy 1744 ca 1745
Db 1641 CA 1642
```

```

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
```

FILING DATE: 29,768
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 30472/114 IMM
REFERENCE/DOCKET NUMBER:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ptzgt-Fls
US-08-232-463-14

Query Match 1.88; Score 45.2; DB 1; Length 7218;
Best Local Similarity 4.08; Pred. No. 0.0053;
Matches 14; Conservative 194; Mismatches 142; Indels 0; Gaps 0;
Qy 1365 ttacacccctgggagcgaagaagaccatgcccctgtgttgcattctctctctgca 1424
Db 1039 TTGGCTCAGGTCAGGAGCTGCCGATYYYYYYYYYYYYYYYYYYYYYY 1098
Qy 1425 gctcctgtgttattctgttggagaatgtgtcactctgttactcttactgcaa 1484
Db 1099 YY 1158
Qy 1485 gacgttattctggaggcttcaagcgcatagtttacacactgaggtctacccca 1544
Db 1159 YY 1218
Qy 1545 cggcaacgcggcctgcctggcaccctgcagcgcatagttcaacactgaggtctctca 1604
Db 1219 YY 1278
Qy 1605 tcactcgttcatcgccaggtgatgctggaatcctctgttacctgactctgagcattt 1664
Db 1279 YY 1338
Qy 1665 acagtggcgtcctcctgctgcccctggcctcctcttcttggccatt 1714
Db 1339 YY 1388

RESULT 4
US-08-501-572-6
Sequence 6, Application US/08501572
Patent No. 6063623
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
TITLE OF INVENTION: DNA sequences encoding it and their use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501,572
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohy, Kimberlin M.
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1896 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-501-572-6

Query Match 1.88; Score 45; DB 3; Length 1896;
Best Local Similarity 47.9%; Pred. No. 0.0028;
Matches 194; Conservative 0; Mismatches 205; Indels 6; Gaps 2;
Qy 1290 gtaaggaggattacgactgctgtggaccacccctctgagtttccaggtgctctg 1349
Db 1362 GTGACAAATATCTACTGGATTTCTTCTACTCTGCCGTGGTGAATTCACGCTGCTCA 1321
Qy 1350 tgactctgtgattatgaccgcctggggcgcaagaagaccatggc-----cctgtgctt 1404
Db 1322 TGATCATCTCATTTATCGACCGCATCGACGCCGTACCCCTGGGCTGCATCAAAATATGG 1381
Qy 1405 tgcatctctctctgagcctcctgcttcttctctctctctctctctctctctctctct 1463
Db 1382 TTGCAGGGGAGCGCTGTGGCCCTCAGTTTTTATACCTGGTGAATCAATATGGCTAAAAA 1441
Qy 1464 ctctgttactctctctgcaagagcgtttattcttggaggtcttcaagcgcatactgtt 1523
Db 1442 TTATTATCTCATGCTTGGGAAGATGGGATCAAAATGGCCATATGAGATAGTCTGCCCTGG 1501
Qy 1524 acacactgaggtctaccccaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1583
Db 1502 TCAATGCTGAGCTGTACCCACATTCATTAGGAATCTTGGGCTCCACATCTGTTCCTCAA 1561
Qy 1584 tggcaagagtggtgct 1643
Db 1562 TGCTGACATGCTGGGATCATACGCGCATTCCTGGTCTACCGGCTCACTAATCTGCGC 1621
Qy 1644 tgtactgactctggcaggtttacagtggtgctgctcctcctggtgctg 1688
Db 1622 TTGAGCTCCCGCTGATGTTTTTCGGCGTACTTGGCTTGGTGGCTG 1666

RESULT 5
US-09-040-444-6
Sequence 6, Application US/09040444
Patent No. 6063766
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk
APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
TITLE OF INVENTION: DNA sequences encoding it and their use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan,Henderson,Parabow,Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040.444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1896 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-040-444-6

Query Match 1.8%; Score 45; DB 3; Length 1896;

Best Local Similarity 47.9%; Pred. No. 0.0028;

Matches 194; Conservative 0; Mismatches 205; Indels 6; Gaps 2;

QY 1290 gtgaggagattacatggacttgctggtgagaccacccctctctgagttccaggttgccttg 1349
DB 1262 GTGACAAATATCTACCTGGATTTCTTCTACTCTGCCCTGGTTGAATCCAGCTGCTTCA 1321
QY 1350 tgactctgtgattatgaccgcctggggcgcaagaagaccatggc-----cctgtgctt 1404
DB 1322 TGATCATCTCATCTTATCGACCGCATCGGACGCGCTTACCCCTGGGCTGCATCAAAATG 1381
QY 1405 tgcattctctctctcagactcctgctgtttatct-gtcttgagaagaatgctca 1463
DB 1382 TTGAGGGGACGCTCTGGCTCTAGTTTTTATACCTGGTGTATCAATGGCTGCTTCAAAA 1441
QY 1464 ctctgttactcttcattgcaagagcgtttatttctggaggcgtttcaagcggcatatgtt 1523
DB 1442 TTATTATCTCATCTTGGGAAGAAATGGGATCACAATGGCCATGAGATAGTCTGCTGG 1501
QY 1524 acacactgaggtctaccaccaggaacgcgggcctcgccctgggcaactgcagcgca 1583
DB 1502 TCAATCTGAGCTGATCCCAATTCATTAGGAATCTTGGCTGCCACATCTCTCTCTCAA 1561
QY 1584 tggcaagagtggtgctctcatcactcccttcctatgcgccaggtgatgctgggaatcctg 1643
DB 1562 TGTGTGACATTTGGTGCATCATCACGCCATTCCTGTCTACCGGCTCACTAATCATCTGC 1621
QY 1644 tgtacctgactctggcagtttaccagtgctgctgctcctcctgctg 1688
DB 1622 TTGAGCTCCCGTATGGTGTCTTGGCGGTACTTGGCTGTGCTG 1666

RESULT 6

US-08-232-463-14/C

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgt-Fls
US-08-232-463-14

Query Match 1.6%; Score 42; DB 1; Length 7218;

Best Local Similarity 4.1%; Pred. No. 0.048;

Matches 15; Conservative 200; Mismatches 155; Indels 0; Gaps 0;

QY 146 aagggccatgagaaggtcctcctgagcgagacatggagagagacttattccag 205
DB 1411 RRR 1352
QY 206 ctaaggcagctccggttgtgaaattccctgcacagcgagagtgcaaggtcagaggac 265
DB 1351 RRR 1292
QY 266 gacacggcttcaggagagcatgaagtcagattgaaggggtccacgtggccttagagct 325
DB 1291 RRR 1232
QY 326 gggagctgagatggggcagctgtgccaaaggatttgccaatcccacatgatact 385
DB 1231 RRR 1172
QY 386 ttcatggtggaagatgcagtggaagccattggctttggaaaatttcagtggaagctgtct 445
DB 1171 RRR 1112
QY 446 gtctcactggttggttggtgctgctgagtcgcatgagatgatgatcctcagcatcctg 505
DB 1111 RRR 1052
QY 506 gcacacacagc 515
DB 1051 GACCTGCACC 1042

RESULT 7

US-08-501-572-4
; Sequence 4, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And\or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/501.572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohy, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-501-572-4

Query Match 1.5%; Score 38.6; DB 3; Length 1882;
Best Local Similarity 47.0%; Pred. No. 0.22;
Matches 224; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

OY 1301 tacatggacttctgtggaccaccctctctgagttccaggtgtctctgtgactctgtgg 1360
DB 1166 TACCTGGACTCTTTTATTCTCTCTGGTGAATTCCTCCCGGCTTCATCATCTGTGC 1225
OY 1361 attatgaccgcctggggcgcaagaagaccat---ggccctgtgttctgtctctctcc 1417
DB 1226 ACCATTGACCGCATTTGGCGCATCTACCCAATAGCGGCTCGAATCTGTGACGGGGCA 1285
OY 1418 ttctcagccctctgtcttctctgtgttggaagaatgtctcactctgttact---c 1474
DB 1286 GCCTGCCCTCCTCATGATCTTTATCCCGCATGAGCTGCACCTGGTTGAACGTTACCTCGCC 1345
OY 1475 ttcatgcaagcgcttattctctgtgaggtcttcaagcgccatgtttcacacactgag 1534
DB 1346 TGCTTGGCCGTATGGGGCCACCAATTGTGCTGCAGATGGTCTGCTGGTGAACCTGAG 1405
OY 1535 gtctaccacacggcaacggccctcgccctggccactcagcgcatggcaagaatg 1594
DB 1406 CHTGACCTTACATTCATCAGGAATCTTGGGATGATGGATGCTCTGCCCTGTGTGACCTG 1465
OY 1595 ggtgtctctcactccgtttcctatcgcccaaggtgatgctgggaatcctctgtgtacctgact 1654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 1466 GGTGGATCTTCACCCCTTCATGCTGTTCAGGCTGATGGAAGTTTGGCAAGCCCTGCC 1525
OY 1655 ctggcagtttacagtggctgtgctctgctgctgctgctgctgctgctgctgttttgcatt 1714
DB 1526 CT---CATTTTGGTGGGTTTGGGCTGACTGTGGGGCATGACTCTTCTTCTCCCA 1582
OY 1715 gagacaaaagcgagagactcagagagtcagagagccagccagggagtgggccagagatg 1771
DB 1583 GAGACCAAGGGTGTGCTTTCCTGAGACTATTGAAGAAGCAGAGAACCTGGGGAGG 1639

RESULT 8

US-09-040-444-4
; Sequence 4, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and\or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Finnegan,Henderson,Farabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-4

Query Match 1.5%; Score 38.6; DB 3; Length 1882;
Best Local Similarity 47.0%; Pred. No. 0.22;
Matches 224; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

OY 1301 tacatggacttctgtggaccaccctctctgagttccaggtgtctctgtgactctgtgg 1360
DB 1166 TACCTGGACTCTTTTATTCTCTCTGGTGAATTCCTCCCGGCTTCATCATCTGTGC 1225
OY 1361 attatgaccgcctggggcgcaagaagaccat---ggccctgtgttctgtctctctcc 1417
DB 1226 ACCATTGACCGCATTTGGCGCATCTACCCAATAGCGGCTCGAATCTGTGACGGGGCA 1285
OY 1418 ttctcagccctctgtcttctctgtgttggaagaatgtctcactctgttact---c 1474
DB 1286 GCCTGCCCTCCTCATGATCTTTATCCCGCATGAGCTGCACCTGGTTGAACGTTACCTCGCC 1345
OY 1475 ttcatgcaagcgcttattctctgtgaggtcttcaagcgccatgtttcacacactgag 1534

```

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;
; TITLE OF INVENTION: From Multiple Transfected Episodes
;
; FILE REFERENCE: 0867/ID903US1
;
; CURRENT APPLICATION NUMBER: US/09/130.114
;
; CURRENT FILING DATE: 1998-08-06
;
; NUMBER OF SEQ ID NOS: 36
;
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 2
;
; LENGTH: 1931
;
; TYPE: DNA
;
; ORGANISM: EBNA
;
; US-09-130-114-2
;

```

	Query Match	1.4%	Score 36.6;	DB 2;	Length 1931;
	Best Local Similarity	46.1%;	Pred. No.	0.88;	
Matches	123; Conservative	0;	Mismatches	144;	Indels 0; Gaps 0;
Oy	146	aagggccatgagaagggtctacctggcgcagcgcgcacatgagaggagacttattccag	205		
Db	1078	AAAGTCAGGAGATGGGGCCCGCAGGTGATGAGGAGCTGGGCCCGCAGGTGATGGAG	1019		
Oy	206	ctaaggaagctgcccggtttgtaaaattccgttcgcacagccagagtgcaaggttcagagagac	265		
Db	1018	GAGTGGGGCCGAGGTGACGGAGAGCTGGGCCCGGAGGTGGAGGCAGGACGGGGAG	959		
Oy	266	gacacggcttcagagagcatgaagtccagattgaagggttcccagtyggccctagaagccl	325		
Db	958	GACGAGGACGGGAGGAGGACGAGGACGGGGAGGACGGGAGGACGAGGACGGGAGGAC	899		
Oy	326	gtgagagctggatgatggggaagctgtcccgaagatttgcctaaccaccagctgatact	385		
Db	898	GGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGAC	839		
Oy	386	tll..Lggtggaagatgcagtgaaagcc	412		
Db	838	CAGGACGGGGAGGACCGGGAGGAGGAC	812		

RESULT 11
US-08-964-127-3
; Sequence 3, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; TITLE OF INVENTION: MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,127
; FILING DATE: 06-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crews, Ph.D., L. Lee
; REGISTRATION NUMBER: P-43,567
; REFERENCE/DOCKET NUMBER: 07334/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906

Db	1346	TGTCCTTGGCGGTATGGGGGCCACCAATTGCTGCAGATGGCTGCCTGGTGAACCGTGAG	1405
Qy	1535	gtctaccaccaggaacgcggccctcgccctgggcacctgcagcggcatggcaagagtg	1594
Db	1406	CTGTACCCCTACATTATCAGGAATCTTGGATGATGGATTGCTTGCCTGTGTGACCTG	1465
Qy	1595	ggtgctctctcatcactcgtttcatgcgcccggtgtagctggaactcctctgtgacctgact	1654
Db	1466	GGTGGGATCTTCACCCCTTCATGTTGTCAGGCTGATGGAACTTTGGCAAGCTGCGCC	1525
Qy	1655	ctggcagtttacagtggtcgtgcctctcgtgcctgcctgcctgcctcttttgccacatt	1714
Db	1526	CT---CATTTTGTGTGGGTTTTGGGCTCACTGCTGGGCCATGACTCTCTCTCCCA	1582
Qy	1715	gaacacaaagcgagagactgcaggaagtcacgcggagtggggcccaggagatg	1771
Db	1583	CAGTCAGAGCTGTGCTTGTGCTTGCTGAGACTATTGAGAAGCAGAGAACCTGGGCAGG	1639

```

RESULT          9
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029, 00
; CURRENT APPLICATION NUMBER: US/09/428, 517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120, 254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106, 100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

```

	Query Match	1.5%	Score 37.6;	DB 4:	Length 50937;
	Best Local Similarity	54.3%;	Pred. No. 3;		
	Matches	76;	Conservative	0;	Mismatches
				64;	Indels
					Gaps
					0;
QY	255	ggtcagaggacgacacaggcttcagagagcatgaagctccagatgaaagggtccacgtgg	314		
Db	37483	ggttggacggttcgccggcactcgtcgcgaatacgcgttgatcggtgagcgagcgcacatga	37542		
QY	315	gcctagagcgtgtggagctgtgatgatggggcagctgtgcccagaagagtttgcacatccca	374		
Db	37543	tctgggacgcggtcgagcgtgtagaactcggtcgctgtagccgagagggttggggat-yacg	37602		
QY	375	cagatgactatttcattgggtg	394		
Db	37603	ccgagcgtttgcacacagtq	37622		

RESULT 10
US-09-130-114-2/C
; Sequence 2, Application US/09130114
; Patent No., 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Dama', Bassam B.
; APPLICANT: Robbins, Alan K.

TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1490 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 492...1349
US-08-964-127-3

Query Match 1.4%; Score 36.4; DB 4; Length 1490;
Best Local Similarity 47.1%; Pred. No. 0.86;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 535 cccaagctggcaggtgacattgctgacccctggtggtctttagcagcatgattccagctc 594
|| |||||
Db 776 CTGGGCTGGCAGGTGATCCCTGGAGCAGATCCTCTTATCTTGGGCTTTGGCTCCGGCTA 835
|| |||||
Qy 595 cagctctggggaataatctcagaccagtcagggcaggaacagggctgaagatcagcgt 654
|| |||||
Db 836 CTTGTTCTCTGGTTACCCCGCAGACAGATTTGGCCGTCGCGGATTTGCTGCTGACCTT 895
|| |||||
Qy 655 gctgtgactctgtactatgagcatccttagtgatttgcgcccgtgtatagctgatact 714
|| |||||
Db 896 GGGGCTCTGGGCCCCCTGTGGAGTAGGAGGGGCTGCTGCAGGGCTCCCTCCACAGCGTCT 955
|| |||||
Qy 715 ggtgtccggggcctggtggcctcgggacgagggagttcccccagtcggtgacgctg 772
|| |||||
Db 956 GGGCTCCGATTCCTCTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTACCTG 1013
|| |||||

RESULT 12
US-08-964-127-1
Sequence 1, Application US/08964127
Patent No. 6277565
GENERAL INFORMATION:
APPLICANT: Grandearl, Andrew David John
TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
MOLECULES
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964.127
FILING DATE: 06-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2460 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 498...2057
US-08-964-127-1

Query Match 1.4%; Score 36.4; DB 4; Length 2460;
Best Local Similarity 47.1%; Pred. No. 1.2;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 535 cccaagctggcaggtgacattgctgacccctggtggtctttagcagcatgattccagctc 594
|| |||||
Db 782 CTGGGCTGGCAGGTGATCCCTGGAGCAGATCCTCTTATCTTGGGCTTTGGCTCCGGCTA 841
|| |||||
Qy 595 cagctctggggaataatctcagaccagtcagggcaggaacagggctgaagatcagcgt 654
|| |||||
Db 842 CTTGTTCTCTGGTTACCCCGCAGACAGATTTGGCCGTCGCGGATTTGCTGCTGACCTT 901
|| |||||
Qy 655 gctgtgactctgtactatgagcatccttagtgatttgcgcccgtgtatagctgatact 714
|| |||||
Db 902 GGGGCTGCTGGGCCCCCTGTGGAGTAGGAGGGGCTGCTGCAGGGCTCCTCCACAGCGTCT 961
|| |||||
Qy 715 ggtgtccggggcctggtggcctcgggacgagggagttcccccagtcggtgacgctg 772
|| |||||
Db 962 GGGCTCCGATTCCTCTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTACCTG 1019
|| |||||

RESULT 13
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 7, 2002, 23:30:57 ; Search time 3543.72 Seconds
(without alignments)
7768.866 Million cell updates/sec

Title: US-09-911-667A-1
Perfect score: 2562
Sequence: 1 cccacgcctccggcgtttgt.....tgaggactgagngccaagtq 2562

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST : *
1: em_estfun : *
2: em_esthum : *
3: em_estin : *
4: em_estom : *
5: em_estpl : *
6: em_estba : *
7: em_estro : *
8: em_estov : *
9: em_hic : *
10: gb_estl : *
11: gb_est2 : *
12: gb_hic : *
13: gb_gss : *
14: em_gss_fun : *
15: em_gss_hum : *
16: em_gss_inv : *
17: em_gss_pin : *
18: em_gss_pro : *
19: em_gss_tod : *
20: em_gss_vrt : *
21: em_gss_other : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	941.4	36.7	972	10	AL523335
2	882.2	34.4	1022	10	AL523336
3	736.6	28.8	778	10	AL529154
4	599.6	23.4	642	10	AL563357
5	556.8	21.7	578	10	AL563357
6	553.2	21.6	638	11	AW663941
7	534.6	20.9	596	10	BG703339
8	502	19.6	536	10	AW957713
9	496.4	19.4	570	11	AW956545
10	463.6	18.1	487	10	BT341271
11	423.2	16.5	468	11	AW173250
12	403.8	15.8	662	11	BF775076
					BT394002

13	364.8	14.2	371	10	AA349847
14	358.2	14.0	366	11	T08271
15	323.2	12.6	355	10	AI372624
16	317	12.4	402	10	AA350545
17	301.8	11.8	450	11	BT296940
18	287.6	11.2	308	11	BG056004
19	286.8	11.2	377	10	AW345918
20	279.8	10.9	420	11	BF412623
21	278.4	10.9	941	10	AU067579
22	276.6	10.8	418	10	BE097508
23	259.4	10.1	266	10	AA351693
24	255.4	10.0	258	11	F08438
25	252.4	9.9	299	11	T10079
26	244.8	9.6	289	10	AW18091
27	201.8	7.9	489	11	BG820031
28	193.6	7.6	360	10	BE656418
29	164.8	6.4	742	10	AU220193
30	163.6	6.4	328	11	BT134483
31	142.6	5.6	1417	12	AK003981
32	138	5.4	441	11	T08675
33	129.6	5.1	253	10	BB586972
34	124.6	4.9	686	10	BE590388
35	122.4	4.8	489	13	AZ979001
36	121.2	4.7	275	10	BB593243
37	115.2	4.5	208	10	BB590911
38	104.6	4.1	996	13	CNS010GB
39	104.6	4.1	1047	13	CNS051V9
40	100.6	3.9	300	11	C57850
41	97	3.8	223	10	BB591622
42	95	3.7	283	10	BB592855
43	92.8	3.6	868	13	CNS04275
44	92.2	3.6	300	10	BB266609
45	89.2	3.5	266	10	BB595082

ALIGNMENTS

RESULT 1
AL523335/C

LOCUS
DEFINITION

AL523335 LTI_NFL003_NBC3 Homo sapiens
prime, mRNA sequence.

ACCESSION
AL523335

VERSION
AL523335.1

KEYWORDS
EST.

SOURCE
human.

ORGANISM
Homo sapiens

REFERENCE
AUTHORS

TITLE
Full-length cDNA libraries and normalization

JOURNAL
Unpublished (2001)

COMMENT
Contact: Genoscope

Genoscope Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DC001YN02"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue.type="neuroblastoma cells"

/lab_host="DH10B"

/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(df) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life

AL523335 972 bp mRNA EST 13-FEB-2001
AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 3

prime, mRNA sequence.

AL523335

AL523335.1 GI:12786828

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 972)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DC001YN02"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue.type="neuroblastoma cells"

/lab_host="DH10B"

/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(df) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 236 a 283 c 252 g 194 t 7 others
ORIGIN

Query Match 36.7%; Score 941.4; DB 10; Length 972;
Best Local Similarity 98.3%; Pred. No. 8.8e-232;
Matches 956; Conservative 4; Mismatches 12; Indels 1; Gaps 1;

Qy 951 tctgctcctgaagtcgaagtgatgtgctgcagggaaccaggaagcaatcg 1010
Db 972 tctggctgccgtgaagtgcaaggtatgctgtcaggaaccaggaataa-gcaatcg 914

Qy 1011 ccaccttaaaagagatagaactaaacaggagagctcccatccgctggggaactcatca 1070
Db 913 CCACCTTAAAGAGGATAGCAACTGAAACGGGGTCCCATGCCGTGGGAAACTCATCW 854

Qy 1071 tctccagacaggaacagaggaacaaatgaggagcctttcacaccccatatttagatgga 1130
Db 853 TCTCCAGACAGAGACCGAGGCAAAATGAGGACCTTCTCACACCCCATTTTATAGTGA 794

Qy 1131 caactttgctgctggtttatatgtgtttccaatgcattctcttaactacgggttagttc 1190
Db 793 CAACCTTYGCTGCTGGTATATATGTTTCCAAATGCATTTCTTACTACGGGTTAGTTC 734

Qy 1191 tactcacacagaactctccagacagagatgctcggcatctccagtcggaagaag 1250
Db 733 TACTCACACAGMACTCTTCAGGACAGAGATGCTCGGGGCTCTCCAGTCGGAAGAAGG 674

Qy 1251 ctgtagagcgaataatgcagctgcctgcgagtagctgagtgagtgagagattacatggact 1310
Db 673 CTGTAGAGGCAAAATGCAGCGTGCCTGCGAGTACCTCAGTGAGGAGATACATGGACT 614

Qy 1311 tgctgtggaacacccctctgaagttccagtgctcctgtgactctgtggattatggacc 1370
Db 613 TGCTGTGGACCACTCTCTGAGTTTCCAGGTGTCCTGTGACTCTGTGGATTTATGACC 554

Qy 1371 gcttggggcgaagaagacacatgccctgtgcttctgcatctctctcttgcagctcc 1430
Db 553 GCCTGGGGCGCAAGAACCAATGCCCTGTGCTTTTGTCATCTTCTCTCTGTCGACCTCC 494

Qy 1431 tgctgtttatctgtgttggagaataatgctcactctgttactcttcaattgcaagagcgt 1490
Db 493 TGCTGTTTATCTGTGTTGGAAGAAATGTGCTTACTCTGTTTACTCTTCATTCAGAGCGT 434

Qy 1491 ttattctggaggttttcaagcggcatatgtttacacacctgaggtctacccccggcaa 1550
Db 433 TTATTTCTGGAGGCTTTTCAAGCGGCATATGTTTACACACCTGAGGTCTACCCACGGCAA 374

Qy 1551 cgcgggcccctgcctggcaccctgcagcggcatggcaagtggtggtgctctcatcactc 1610
Db 373 CGCGGGCCCTCGGCTGGGCACTTCGACGGCATGGCAAGAGTGGGTGCTCTCANTACTC 314

Qy 1611 cgttcaatgccaggtgatgctggaatctctctgtgtacctgactctgtgagctttacagtg 1670
Db 313 CGTTATCGCCACAGTGATGCTGGAATCCTCTGTGTACTCTGACTCTGGCAGTTTACAGTG 254

Qy 1671 gctgctcctctggtccttagctccctgctcttttttggccattgagacaaagcggag 1730
Db 253 GCTGCTGCCTCCTGGCTGGCCCTGCTCTGCTTTTGGCCATTGAGACCAAGAGCCGAG 194

Qy 1731 gactgcaggaggtccagccacagggagtgggccaggagatggctgcggcggaggaatgcacg 1790
Db 193 CACTGCAAGAGTCCACGCCACCGGAGTGGGGCCAGGAGATGCTCGCGCGAGGATGCAGC 134

Qy 1791 gtcaggtgttacaggtcgaactctggtctctcaaggaatagtcaccgatggggagactgag 1850
Db 133 GTGCAGGTGTTTACAGGTGCAACTCTGGCTCTCAGGAATAGTGACCGATGGGGGACTGAG 74

Qy 1851 ctggtcttttgaggtgcagagcttgggggctggtgcagggcccaactggggcactgattgt 1910

Db 73 CTGGTCTTTGAGCTGCAGAGCTTGGGGGCTGGGAGGCCCGCAGCTGGGGCACTGATTGT 14

Qy 1911 cactgcccacatc 1923
Db 13 TACTCCCGACATC 1

RESULT 2
AL523336 1022 bp mRNA EST 13-FEB-2001
LOCUS AL523336 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 5
DEFINITION prime, mRNA sequence.
ACCESSION AL523336
VERSION AL523336.1 GI:12786829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YN02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 180 a 268 c 326 g 243 t 5 others
ORIGIN

Query Match 34.4%; Score 882.2; DB 10; Length 1022;
Best Local Similarity 98.0%; Pred. No. 1.7e-216;
Matches 952; Conservative 4; Mismatches 7; Indels 8; Gaps 6;

Qy 13 ggccttctcgtgggtggtcccccagcgctccctcccccaaacagctgctgcacagg 72
Db 59 GCCTTTGTCTGCTGGCTGGTCCCGCAGCGGTCTCCCTCCGCAACAGCTGCTCCAGGG 118

Qy 73 aggaagcggcggg-gtgcctcagacttcccgggtgctgaaacggagggtcgtcactc 131
Db 119 AGGAAGCGGCGGNGTCTGTCTCAGCTTCCCGGTGCTGAAACCGGAGGGCTCTGTCATC 178

Qy 132 caccactaccatgtaagggccatgagaagggtctatcctcgtcgcagcgcggacatggagg 191
Db 179 CACCACATACCATGTAAAGGCCATGAGAAAGGCTCATCTCTGCGCA-SGCGGACATGGAGG 237

Qy 192 aggacttattccagtaaggcagctgcggtgtgtgaaattccgtgcacagcggcagagt 251
Db 238 AGGACTTATTCCAGCTAAGGCAGCTGCCGTTGTGTAATTCCTCGCACAGCGGAGATG 297

Qy 252 caaggtcagagacacacgcttcagagagcatgaagtccagattgaaggggtccacg 311

```
Db 298 CAAGGTCAGAGGACGACACGCGCTTCAGGAGAGCATGAAGTCCAGATTGAWGGGTCCACG 357
Qy 312 tggcgctagagcgtgtgagctgagatgagtgagggcagctgtgcccagagagtttcccaatc 371
Db 338 TGGGCTAGAGGCTGTGGAGCTGGATGATGGGCGAGCTGTGCCAAGAGTTTGCCAAATC 417
Qy 372 ccaccygtatgactttcatgtggaagatgcagtggaagccatggctttggaataatc 431
Db 418 CCACCGATGATACCTTCATGTGTGAAGATGCAGTGGAGCCATTGGCTTTGGAAAAATTC 477
Qy 432 agtgaagctgtctgttctcactgctgtgcttgatgctgagtgatgagtgatgag 491
Db 478 AGTGAAGCTCTCTCTCTCAGCTGGCTTGGCTTGGATGGCTGATGCCATGAGATGATGA 537
Qy 492 tctcagatcctgagcaccagctgcatgagtgagtgagtgagtgagtgagtgagtgag 551
Db 538 TCTCAGATCCTGGCACCAGCTGATGCGAGTGGAGGCT--CCAAGCTGGCAGGTGG 596
Qy 552 cattgtgacctggtgtgttctttagtgatgagtgatgagtgagtgagtgagtgagtgag 611
Db 597 CATTGCTGACCTCGTGTCTTTCTAGGATGATGTCACGCTCCACGCTCGGGGAAATA 656
Qy 612 tctcagaccagtcagcaggaacacagggctgaagatcagctgctgagctgtgactctgact 671
Db 657 TCTCAGACCACTACCGGAGGAAACAGGGCTGAAGATCAGCGTCTGTGGACTCTGTACT 716
Qy 672 atgacatctttagtgcatttgcgcccgtgtatagctgagtcctgagtgagtgagtgag 731
Db 717 ATGCGATCCTTAGTGATTTGCGCCCGTGTATAGCTGGATCTCTGCTGCTGCTGCTGCT 776
Qy 732 tggcttcggagtcgagggagttcccccagtcggtgagcgtgtatgagcgtgtatgagcgt 791
Db 777 TGGGCTTCGGATCGGAGGAGTTCGCCAGTCGGTGACGCTGTATGCGGAGTTCCTTCCCA 836
Qy 792 tgaagccagagctaaatgtatttctgagtgagtgagtgagtgagtgagtgagtgagtgag 851
Db 837 TGAWAGCCAGAGCTAAATGTATTTTGTGTGATTTGAGGTATTTCTGGGCGCATCGGGACAGTGT 896
Qy 852 tggaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 911
Db 897 TCGAGGTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955
Qy 912 tctcagctgtccgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 971
Db 956 TCTCAGCTGTCC--GCTCCCTCTCTTTGCCGTGTGTGTTCTGCTGCTGCTGCTGCTGCTGCTG 1011
Qy 972 ggtatgagtg 982
Db 1012 GGTATGATGTG 1022
```

RESULT 3

```
AL529154
LOCUS AL529154 778 bp mRNA EST 13-FEB-2001
DEFINITION AL529154 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 5
prime, mRNA sequence.
ACCESSION AL529154
VERSION AL529154.1 GI:12792647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 778)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .778
```

FEATURES

SOURCE

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
```

BASE COUNT 150 a 193 c 254 g 166 t 15 others
ORIGIN

Query Match 28.8%; Score 736.6; DB 10; Length 778;
Best Local Similarity 97.8%; Pred. No. 5.1e-179;
Matches 752; Conservative 13; Mismatches 2; Indels 2; Gaps 2;

```
Qy 13 ggctttgtctgtggctgtcccccagcggctcctccccgaacagctgctgctccagg 72
Db 11 GGCTTTGTCTGTGGGTGGTCCCGACGGCTCCCTCCCGAACAGCTGCTGCTCCAGGG 70
Qy 73 aggaagcggcgcggtgctgtccagcttcccggtgctgaaacccgagggctgctcatcc 132
Db 71 AGGAAGCGGCGGGTGTCTGTCCAGCTTCCCGGTGCTGTAACACCGGAGGGCTGTCATCC 130
Qy 133 accactaccatgttaagggccatgagaagggctcatctggcgagcgcgacatggagga 192
Db 131 ACCACTACCATGTAAAGGCCATGAGAAGGGCTCATCTCTGGCGCA-CGCGGACATGGAGGA 189
Qy 193 ggacttattccagctaaagcagctgccgg-ttgtgaaattccgtgcgacagcgagagtg 251
Db 190 GGACTTATTCCAGCTAAGGCAGCCCGCGGTTGTGAATAWTCCTGCGCACAGCGGAGAGTG 249
Qy 252 caaggtcagaggaagcagcagcgttcagagagcagtgagtgagtgagtgagtgagtgag 311
Db 250 CAAGGTCAAGAGGACGACACGGCTTCAGGAGAGCATGAAGTCCAGATTGAAGGGGTCCACG 309
Qy 312 tgggcttagagggctgtgagcgtgagtgatggggcagctgtgcccagagagttgccaatc 371
Db 310 TGGGCTTAGAGGCTGTGGAGCTGGATGATGGGGGAGCTGTGCCCAAGAGAGTTTGCCAATC 369
Qy 372 ccaccgatgatacttcatgtggaagatgcagtggaagccatggctttggaataatc 431
Db 370 CCACCGATGATACCTTCATGTGTGAAGATGCAGTGGAGGCCATTGGCTTTGGAAAAATTC 429
Qy 432 agtgaagcgtgtgttctcactggttggcttggttgagtggtgagtgagtgagtgagtgag 491
Db 430 AGTGAAGAGCTGTCTGTCTCAGTGGCTTGGCTTGGATGGCTGATGCCATGAGATGATGA 489
Qy 492 tctcagcatccttgagcaccagcagctgcatgagtgagtgagtgagtgagtgagtgagtgag 551
Db 490 TCCTCAGCATCTCGGCCACACAGCTGCTGCGAGTGGAGGCTCCCAAGCTGGCAGGTGG 549
Qy 552 cattgctgacctggtgtgttctttagtgatgagtgatgagtgagtgagtgagtgagtgag 611
Db 550 CATTGCTGAMCTGGTGGTCTTTKTAGCATGAGTCCAGCTCCAGCTCTCGGGGAATA 609
Qy 612 tctcagaccagtcagcaggaacacagggctgaagatcagctgagtgagtgagtgagtgagtgag 671
Db 610 TTTCAAGACAGTACGGCAGGAAACARGGCTGAAGATCAGCGTCTGCTGCTGCTGCTGCTGCT 669
Qy 672 atggaatccttagtgatcttggcccggtgtatagctggatcctggtgctcggggcctgg 731
Db 670 ATGGCATCCCTTAGTGCAATTTGCCGCCGCTKTWTAGCTGGATCCTGGTGTCTCCGGGGCCTGG 729
```

```

QY 732 tgggttcgggagtcgagagtgctccacgtcggtgacgtgtatgccga 780
|||||
Db 730 TGGGCTCGGATCGGAGGAGTTCGCCAGTTCGGTACGCTKTTCGCGA 778

RESULT 4
AL563357 642 bp mRNA EST 16-FEB-2001
LOCUS AL563357.LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563357
VERSION AL563357.1 GI:12912671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 149 a 192 c 170 g 117 t 14 others
ORIGIN

Query Match 23.4%; Score 599.6; DB 10; Length 642;
Best Local Similarity 96.0%; Pred. No. 9.9e-144;
Matches 616; Conservative 2; Mismatches 23; Indels 1; Gaps 1;

QY 1221 atgtctgcggcatctccagtcggaagagctgttagaggcaaaatgcagctggcctgcg 1280
|||||
Db 642 ATTTCTGGCGCTTCTCCAGTCGGAARAAGGCTGAAGAGGCAAAATGCAGCTGGCTGCG 583

QY 1281 aatacctgagtgagagattacatggactgtctgtgaccacacctctctagttccag 1340
|||||
Db 582 AGTACTGAGTGAGGAGATWACATGGACTTGTCTGTGACCACCCCTCTCTGAGTTTCCAG 523

QY 1341 gtgtcctgtgactctgtgattattgaccgctggcgcaagaagaccatggccctgt 1400
|||||
Db 522 GTGTCTGTGACTGTGTGATTATTGACCGCTGGGCGCGCAAGAACCATGGCCCTGG 463

QY 1401 gctttgcatctctctctctgcagcctctgtgtttattctgtgttggaag-aaatgtg 1459
|||||
Db 462 GCTTTGTCTATCTCTCTCTTCGACGCCNCCNGCTGTTNNCTGTGTGGAAGAAATGNG 403

QY 1460 ctactctgttactcttcattgcaagacgtttattcttgaggctttcaagcggcatat 1519
|||||
Db 402 CTCACCTGTACTCTTCATGGCAAGACGGTATTCTTGAGAGGCTTTCAAGCGGCATAT 343

QY 1520 gtttacacacctgaggtgtactaccacacgcgagcgccctcgccctgggcacctgcagc 1579
|||||

```

```

Db 342 GTTNACACACCTGAGGTCTACCCACGCAACGCGGCCCTCGCTGGCACCTGCAGC 283
|||||
QY 1580 ggcattgcaagagtggtgtctctctcatcactcctgtcatcgcacaggtgatgctggaatcc 1639
|||||
Db 282 GGCATGGCAAGAGNGGGGCTCTCANTACTCCGTTTCATGCCAGGTTGCTTGGAAATCC 223
|||||
QY 1640 tctgtgactgactctgcagtttacagtggctgctgctcctgctgctcctgctgctcc 1699
|||||
Db 222 TCTGTGACTGACTCTGCAGTTAACAGTGGCNGCTGCTCTGCTGCTGCTGCTGCTGCTCC 163
|||||
QY 1700 tgccttttgcctattgagacaaagcgagagtagtccagagtcacagcaccacggagtg 1759
|||||
Db 162 TGCCTTTTGCCTATTGAGACCAAGGCCGAGNACTCCAGAGTCCAGCACCACGGAGTGG 103
|||||
QY 1760 gcccaggagatggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1819
|||||
Db 102 GGCCAGGAGATGGTCGGCCGAGGAATGCACGGTGCAGGCTGTACAGGCTGGACTCGGC 43
|||||
QY 1820 tctcaggaatagtgaccatggggactgagctgagctggtctttga 1861
|||||
Db 42 TCTCAGGAATAGTACCGATGGGGGACTGNGCTGGTCTTTGA 1
|||||

```

```

RESULT 5
LOCUS AW663941/c
DEFINITION hi73h10_x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2977987 3' similar to TR:Q9Z17 Q9Z17 SV2 RELATED PROTEIN.
; mRNA sequence.
ACCESSION AW663941
VERSION AW663941.1 GI:7456480
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 578)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
Location/Qualifiers
1..578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2977987"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 137 a 170 c 159 g 111 t 1 others
ORIGIN

```

FEATURES
source

Location/Qualifiers
1..578

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2977987"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 170 c 159 g 111 t 1 others
ORIGIN

Query Match 21.7%; Score 556.8; DB 10; Length 578;
Best Local Similarity 99.3%; Pred. No. 1e-132;

	Matches	569;	Conservative	0;	Indels	3;	Gaps	1;
QY	1377	ggcgcaagaagacat-ggcccgtgtcttgctcattcttctcgcagccctcgtctg	1435					
Df	578	GGCCAAAAGACCATGGGCCCTGTGTCATCTTCTCCTTCTGCAGCCTCCTGCTG	519					
QY	1436	ttaactgtgtlgaagaataatgtgctcactctgttaactttcatgcaagagcgtttati	1495					
Df	518	TTTATCTGTCTTGGAAGANATGCTCACTCTGTATTACTTTCATTTGCAAGAGCGTTTATT	459					
QY	1496	tctgaaggcttcaagcgcgatattttaacacctaaagtctaccacacgcgaacgcgg	1555					
Df	458	TCTGAGGCTTTCAAGCGGCAATGTTTACACACTGAGGTCTACCCACCAGCACGCG	399					
QY	1556	gccctcgccctggcaacctgcagcgcgcgaagagtggtctctcatcactccgttc	1615					
Df	398	GCCCTCGGCTGGCACCTGCAGCGGATGCAAGAGTGGGTGCTCTCATCACTCCGTTC	339					
QY	1616	atcgcccaagtgtatgtggaatctctgttacctgactctggcagttacagttagctgc	1675					
Df	338	ATCGCCCAAGTGTATGTGAATCTCTGTGTACCTGTGGCAGTTTACAGTGGCTGC	279					
QY	1676	tgcctctggtcgccctggcctcctgcttttggccattgagaccaagcgcgcgcgtctg	1735					
Df	278	TGCCTCTGGCTGCCCTGCTCTGCTTTTTGCCATTGAGACCAAAGCGAGGACTG	219					
QY	1736	caggagtcacaccacccgggaatggggccagagatgctcggccgaggaatgcacagtgca	1795					
Df	218	CAGGAGTCTCAACCCACCGGAGTGGGGCCAGAGATGTGTGGCCGAGGAATGCCACGGTCA	159					
QY	1796	gggtaccagagtcgaactctgctctcagaaatagtgaccgatggggagctgagctggt	1855					
Df	158	GGTGTACCAGTCTGAATCTGGCTCTCAGGAATAGTAGCCGATGGGGGACTGACTGGT	99					
QY	1856	cttgaagctgcagagcttgggggctggcagcccccaactggggcactgattgtcaatg	1915					
Df	98	CTTTGAGGCTGCAGAGCTTGGGGGCTGGCAGGCCCACTGGGGCACTGATTGTCACTG	39					
QY	1916	ccgacatcaagaactcacccaagatgacct	1948					
Df	38	CCGACATCAAGAATCAACCAAGAGTAGACCT	6					
RESULT	6							
LOCUS	BG703339	638 bp	mRNA	EST	07-MAY-2001			
DEFINITION	60268528F1 NIH_MGC_95	Homo sapiens cDNA clone IMAGE:4817829	5'					
ACCESSION	BG703339							
VERSION	BG703339.1	GI:13975571						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.							
JOURNAL	1 (bases 1 to 638)							
COMMENT	NH-BMC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: csapbs-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LIA010718 row: 1 column: 22 High quality sequence stop: 638. Location/Qualifiers 1..638							
FEATURES								

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817829"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH MGC library."
123 149.6 212.9 144.1
BASE COUNT

```

**BASE COUNT
ORIGIN**

Query Match 21.6%; Score 553.2; DB 11; Length 638;
Best Local Similarity 97.8%; Pred. No. 9.1e-132;
Matches 624. Conservative 0; Mismatches 8; Indels 6;

162	Qy	gctcatcctggcgagcgcgacatggagggaggaactattccagctcaaggcagctgcggg	221
1	Db		
2	Db	gcgagcttggggcgagcgcgacatggagggaggaactattccagctcaaggcagctgcggg	61
222	Qy	tgtgaaatccgtcgcacagggcgagagtga..ggtcagaggagcagacgggttcaggag	281
1	Db		
62	Db	ttgtgaaattccgtcgcacagggcgagagtga..aaggtcagaggagcagacgggttcaggag	121
282	Qy	agcatgaagtcacaatgaaaggggtccacatgggctcagaaggtgtggagctgagatgagtg	341
1	Db		
122	Db	agcatgaagtcacaatgaaaggggtccacatgggctcagaaggtgtggagctgagatgagtg	181
342	Qy	gggcagctgtgtcccaagaggtttggccaatccccacgatgatactttcatgtgtgaaagtg	401
1	Db		
182	Db	gggcagctgtgtcccaagaggtttggccaatccccactgatgatgatactttcatgtgtgaaagtg	241
402	Qy	cagtggaagccatggtgttttgaaa..aatttcagtggaagctgtctgtttcactgcgtctg	460
242	Db	cagtggaagccatggtgttttgaaa..aatttcagtggaagctgtctgtttcactgcgtctg	301
461	Qy	gcttgatggtcgtgcctatggagatgatcctcagcaatcctggcaccacacagctgcgat	520
302	Db	gcttgatggtcgtgcctatggagatgatcctcagcaatcctggcaccacacagctgcgat	361
521	Qy	tgcagtggaagctcccaagctgagcagtgagcattcagaccctcgatggtctttgttaggc	580
362	Db	tgcagtggaagctcccaagctgagcagtgagcattcagaccctcgatggtctttgttaggc	421
581	Qy	atgatgtccagctccacagctctgggaaatatctcagaccagctacggca..ggaaaaacagg	639
422	Db	atgatgtccagctccacagctctgggaaatatctcagaccagctacggca..ggaaaaacagg	481
640	Qy	gctaagaatcagcgtcgtgtgacctctgta..ctatggcattccttagtgcatttgcgcccgc	698
482	Db	gctgcagatcagcgtcgtgtgacctctgta..ctatggcattccttagtgcatttgcgcccgc	540
699	Qy	tgtatagctggaatcctggtgctccggggcctggtgggcttcggagat..cgagaggagtcccc	757
541	Db	tgtatagctggaatcctggtgctccggggcctggtgggcttcggagat..cgagaggagtcccc	600
758	Qy	c..agtcggtgagcgtgtatgctccgagttccttccccatga	794
601	Db	caagtcgctgagcgtgtatgctccgagttccttccccatga	638

RESULT

RESULT	7
AW957713	
LOCUS	
DEFINITION	AW957713 596 bp mRNA
ACCESSION	AW957713
	EST
	01-JUN-2000
	MAGE resesquences, MAGE Homo sapiens cDNA, mRNA sequence.

```
VERSION      AW957713.1  GI:8147396
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
LOCUS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DEFINITION  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 596)
AUTHORS     Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
TITLE       Assessment of gene expression patterns in a model of colon tumor
JOURNAL      metastasis using a 19,200 element cDNA microarray
COMMENT      Unpublished (2000)
            Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 110
            Seq primer: Reverse.
FEATURES     Location/Qualifiers
             Source
              1..596
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="MAGE resequencences, MAGE"
              /note="vector: pBluescriptskm"
BASE COUNT   133 a 170 c 149 g 144 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

```
Db 551 GCCCCCACCCATCCCCCTT--GGAGTTGTCGTTCTGTCGCAACCAAGAGA 596
RESULT      8
AW965645    536 bp  mRNA  EST  01-JUN-2000
LOCUS       EST377718 MAGE resequencences, MAGI Homo sapiens cDNA, mRNA sequence.
DEFINITION  AW965645
ACCESSION   AW965645
VERSION     AW965645.1  GI:8155481
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
LOCUS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DEFINITION  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 536)
AUTHORS     Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
TITLE       Assessment of gene expression patterns in a model of colon tumor
JOURNAL      metastasis using a 19,200 element cDNA microarray
COMMENT      Unpublished (2000)
            Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 217
            Seq primer: Reverse.
FEATURES     Location/Qualifiers
             Source
              1..536
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="MAGE resequencences, MAGI"
              /note="vector: pBluescriptskm"
BASE COUNT   95 a 144 c 174 g 123 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```


Query Match 18.1%; Score 463.6; DB 10; Length 487;
Best Local Similarity 96.9%; Pred. No. 1e-108;
Matches 472; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy	1340	ggtgtcccttgtgaactctgtggaattattgaacgcctcgtggggcgcaagaacacatggccctt	1399
Db	487	GGTGTCTATTGTGACTCTGTGGATTATTGACCGCATGGGGCGCAAGAGACCATGGCCNTG	428
Qy	1400	tgccttggcatctctccctctgcagcctcctcgtcgttattctctgttggaaagaaatgtg	1459
Db	427	TGCTTGCTCATCTCTCCCTTGTGACGCTCGTGTGTTTATCTGTGTGGAAGAGATGTG	368
Qy	1460	ctcaactctgttactcttctaatcgaagacggtttattctcgagagcctttcaagcggcaatat	1519
Db	367	CTCACTCTGTTACTCTTTCAATTCGAAGAGCGTTTATTATTGAGGCTTTCAACCGGCATAT	308
Qy	1520	gtttacacaccttgaggtctacccccagcgcaacgcgggcctctcggcctgggcacctgcagc	1579
Db	307	GTTTACACAGCTGAGGTCTACCCACGCGCAACGCGGGCCCTCGGCCCTGGGCACCTGCAGC	248
Qy	1580	ggcatgucgaagagtgggtgtctctcatcactcgttcatcgcgcaggtgtagctgaaatcc	1639
Db	247	GGCATGCGAAGAGTGGGTGCTCTCATCACTCCGTTCTATCGCCACAGGTGATGCTGGAATCC	188
Qy	1640	tcigtgtacctgactctggcagtttacagtggtcgtcctcctctgctgcctcgtgcctgcctcc	1699
Db	187	TCGTGTCACTTGACTCTGGCAGTTTACAGTGGCTGCTGCCCTCTGGCTGCCCTGGCCTCC	128
Qy	1700	tgccttttgcgccattgagaccaaaggcgaggaactgcaggagtcagaccacccggagtg	1759
Db	127	TGCTTTTGTGCCATTGAGACCAAGAGCCGAGGACTCCAGAGTCCAGCACCCGGGAGTGG	68
Qy	1760	ggccagagagatgctcggcgcaggaatgcacggtgcaggtgttaccaggttcgaactctggc	1819
Db	67	GGCCAGAGAGATGTGCGCCCAAGGAATGCACGGTTCAGGTGTCCCAAGTTCGAACCTCTGCC	8
Qy	1820	tctcagg	1826
Db	7	TCTCAGG	1

```

RESULT, 11
BF775076      468 bp      mRNA      EST      25-APR-2001
LOCUS
DEFINITION   285128 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF775076
VERSION     BF775076.1 GI:12122976
KEYWORDS
SOURCE      EST.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
            Bovidae; Bovinae; Bos.
            1 (bases 1 to 468)
            Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrénkrug,S.C., Bennett
            G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            pörtea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keeler,J.W.
TITLE       Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
            Genome Res. 11 (4), 626-630 (2001)
            21180013
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
            PCR Primers

```

FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTTCCAGTCCAGCAGC
 Plate: 86 row: E column: 5
 Seq primer: ATTTAGGTGACACTATAG.

FEATURES	SOURCE
----------	--------

```

1. 468
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DHI10B"
/note="Vector: pCMV SPORT6;
Library made from pooled testis
macrophage, ovary, fetal se-
longissimus muscle."
87 a 126 c 139 c 116 t
BASE COUNT

```

BASE COUNT	87 a	126 c	139 g	116 t
ORIGIN				
Query Match		16.5%	Score 423.2;	DB 11; Length 468;
Best Local Similarity		94.0%	pred. No. 2.6e-99;	
Matches 440: Conservative		0;	Mismatches 28;	Indels 0; Gaps 0;

QY	671	tatggcatccttagtgcatttgcacccgtatagctgatacctggtctcggggcctg	730
Db	1	TACGGCATCCACAGTGGCTTTGCGCCCGTGTACACTGTGATCCTGGTCTCGGGGCGCTG	60
QY	731	gtgggtctcgggatacggaggagttcccccagtcggtagcgtgtatgcgcaggtctctccc	790
Db	61	GTGGCTTCGGGATCGGAGGGGTCCCCACAGTCGGTACACTGTATGCTCAGTTCCTTCGCC	120
QY	791	atgaaacccagagctaaatgtatttttgcgtattgaggtatttctggccatcgggacagt	850
Db	121	ATGAAGCCAGAGCTTAATGTATTTTGTCTGATTGAGGTCTCTGGCCATCGGGACAGTG	180
QY	851	ttcgaagtcgtcctgctgtgttcgtgatgccacgctgggctggcttggctgtctc	910
Db	181	TTCGAGGTGCTCTGGCTGTGTTTCTGATGCCACGCTGGGCTGGCTGGCTGCTCATC	240
QY	911	ctcagctgtcccnctcctcctcttltgccgtgtgtgttttctgctacctgaagtcca	970
Db	241	CTCTCAGCTGTGCCACTCTCTCTTTTGGGTCTCTGTGTTTGGCTGCCGAGAGTGGC	300
QY	971	aggtatgatgtgtctcagggaaacaggaaaggcaatcgccacttaagaggatagca	1030
Db	301	AGGTATGATGTCTCTCCGGGAACAGGAGAAAGGCCATCGCCACCTTTAAAGAGAAATGCC	360
QY	1031	actgaaacggagctcccatgcgcgtggggaacatcatctccagacagaaaccca	1090
Db	361	ACGGAAAACGGAGCTCCCATGCTCTGGGGAACATCATTTCTTCAGACAGAGAACCGA	420
QY	1091	ggcaaatgaggaccttttcacacccatttagatgacacactttg	1138
Db	421	GGCAAAATCAGGAGCTTTTTCACACCCATTTTATGATGACAACTTG	468

RESULT 12
DT204003

BI394002	BI394002	662 bp	mRNA	EST	06-AUG-2001
LOCUS	pgpin.pk013.c4	Normalized Chicken Pituitary/Hypothalamus/Pineal Library Gallus gallus cDNA clone pgpin.pk013.c4 5', similar to gb AAC78627.1 (AF060173) SV2 related protein [Rattus norvegicus], mRNA sequence.			
DEFINITION	BI394002				
ACCESSION	BI394002				
VERSION	BI394002.1	GI:15087284			
KEYWORDS	EST.				
SOURCE	chicken.				
ORGANISM	Gallus gallus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (bases 1 to 662)				
AUTHORS	Porter,T.E. and Cogburn,L.A.				

Qy	1287	tga	tgagagagattacatgacttgc	tgtgagaccacccctctgagttccagagtgcc	1346
Db	61	TCACAGAGGAAGACTACACTGATCGCTCGACAAACCCGTGTCAGAAATTC	CCAGGTGCT	120	
Qy	1347	ttg	gaactctgtgattattgacgccttggggcgcaagaagaccatgccctgtgctttg	1406	
Db	121	TAGTAACACTCTGGATTATGATCGGATAGGCGCGAAGAAACCATGGCCCTGTCCTTCT	180		
Qy	1407	tcatctctctctctgcagcctcctgcgtgttattctgtgttggagaagaatg	gtgcacac	1466	
Db	181	TGTCTCTCTCATTTTGCAGCCTGCTGCTTTCTCTGTTGGAAAGAAATGT	TCTTACTTG	240	
Qy	1467	tgtactcttcaattgcaagagcggttattcttggaggctttcaagcggcata	tgtttaca	1526	
Db	241	TGCTGCTCTTCATTGCAAGAGCTTTTATTTACAGGAGATTTCA	GCTGCTTATGTTTACA	300	
Qy	1527	caactgaggtctaccccacggaacgcggcgccctggcctgggcaactgcagc	ggcgatgg	1586	
Db	301	CTCCGAGGTTTACCCAAACAGCCACGCGTGCCTGGGCCGTGGGAACAT	CAGTGGGAATGG	360	
Qy	1587	caagagtggtgctctcatcactcgttcatgccacagtgatgctggaa	ctctctgtgt	1646	
Db	361	CCAGAGTGGGAGCCCTCATTAACACCATTCATTG	CACAGGTGATTTGGAAATC	420	
Qy	1647	acctgactctggcagtttacagtggctgtgcctctggctggcctggcctc	tgctgtttt	1706	
Db	421	ATTTAACACTGGCAGTTTACAGTGGATGTGCTGTGCTGGCTGCTGTG	GGCTTCTTCTTCT	480	
Qy	1707	tgccattgagaccaaagcgagaaactcaggaggtccagccacgggagtg	ggggccagg	1766	
Db	481	TGCCCCATTGAACAAAGGTCTGGCTTACAGAGGTCCAGGCCACAGAGAAT	GGGACAGG	540	
Qy	1767	agatgctgcgcgaggaatgcacggtgcaggtgttaccagggtgc	aaactctggctcagg	1826	
Db	541	AGATGCTGGGAGAGATCTCATTTCCCGAGAGTACCAGGTCTAACT	TGGGTACAGG	600	

Query Match	14.28;	Score 364.8;	DB 10;	Length 371;
Best Local Similarity	98.7%;	Pred. No. 2.7e-83;		
Matches 366;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
QY 1169	ttctcttactacggttagttctactaccacagaactcttccaggcaggagatgctgc	1228		
Db 1	TTCTCTTACTACGGCTTAGTTCTACTCCACAGAACTCTTCAGCGCAGGAGATGCTGCG	60		
QY 1229	gcatctccagtcggaagaaggctataagagcaaaatcagcctggcctggagtagctcg	1288		
Db 61	GGCATCTCCAGTCGGGAAGGCTGTGAGGCAAAATGCAGCGCTGGCCTGGCAGTAGACCTG	120		
QY 1289	agtggaggagattacatggacttcgttggaccacctctctctgagttctccagggtcctt	1348		
Db 121	AGTNAGNGGATTACATGGACTTGTGTGGACACACCTCTCTGAGATTTCACAGGTGCTTT	180		
QY 1349	gtgaactctgaaattatgaccgcctggggcgcaagaagaccatggccctgtgctttgctc	1408		
Db 181	GTGACTCTTWTGATATTACCGCCTGGGCGCGCAAGAGACCATTGGCCCTGTGCTTGTGC	240		
QY 1409	attctctctctgcagccctcctgctgttttatctgtgttggagaagaattgctcactctg	1468		
Db 241	ATCTCTCTCTTCTGCAGCCTCTGCTGTATTATCTGTGTTGGAAGAAATGCTCACTCTG	300		

[illegible]

Search completed: March 8, 2002, 00:47:58
Job time: 4621 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 03:28:23 ; Search time 5957.55 Seconds
(without alignments)
4552.436 Million cell updates/sec

Title: US-09-911-667A-3
Perfect score: 1644
Sequence: 1 atggaggagactattcca.....cgaactctggctctcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1326.2	80.7	2421	10	AF060173	AF060173 Rattus no
2	432.4	26.3	2104	9	HS0802697	AL359592 Homo sapi
3	209.4	12.7	212183	2	AC087893	AC087893 Homo sapi
c	91	5.5	11856	1	AE007266	AE007266 Sinorhizo
	89.4	5.4	148432	3	AC004642	AC004642 Drosophil
	89.4	5.4	157851	2	AC020509	AC020509 Drosophil
6	89.4	5.4	300542	3	AE003462	AE003462 Drosophil
7	89.4	5.4	198677	1	AE001863	AE001863 Deinococc
8	89	5.1	19834	3	CEK01F9	222175 Caenorhabdi
9	84	5.0	10344	1	AE004767	AE004767 Pseudomon
10	82.8	5.0	10344	1	AE004767	AE004767 Pseudomon
11	81.8	5.0	62204	2	AC067757	AC067757 Homo sapi
12	77	4.7	11286	1	AE008141	AE008141 Agrobacte
13	74.2	4.5	81020	8	AB026645	AB026645 Arabidops
14	74.2	4.5	81020	24	AC024128	AC024128 Arabidops
15	70	4.3	13462	1	AF218267	AF218267 Pseudomon
16	69.6	4.2	2622	10	AF060174	AF060174 Rattus no
17	69.2	4.2	1638	9	AF052188	AF052188 Homo sapi
18	69.2	4.2	4948	9	AB018278	AB018278 Homo sapi
19	66.2	4.0	2315	5	DYGSV2TRAN	L23403 Discopyge o
20	66.2	4.0	3190	9	BC000776	BC000776 Homo sapi
21	66.2	4.0	4353	9	AB018279	AB018279 Homo sapi
22	66	4.0	3660	10	RATSV2BA	L10362 Rattus norv
23	64	3.9	35000	1	AB000617	AB000617 Bacillus
24	64	3.9	221160	1	BSUB0002	299105 Bacillus su
25	63	3.8	2268	6	AX079057	AX079057 Sequence
26	62.4	3.8	10029	1	AE007908	AE007908 Agrobacte
27	61.6	3.7	16775	1	AB000735	AB000735 Nocardioid
28	61.6	3.7	40699	3	CEZK637	211115 Caenorhabdi
29	59.8	3.6	7424	1	AB050935	AB050935 Pseudomon
30	59.2	3.6	11886	1	AE004638	AE004638 Pseudomon
31	58	3.5	159167	2	DMBR42L18	AL122031 Drosophil
32	57.2	3.5	37221	2	AC012886	AC012886 Drosophil
33	57.2	3.5	126340	2	DMBR18C8	AL122030 Drosophil
34	57.2	3.5	299537	3	AE003438	AE003438 Drosophil
35	56.6	3.4	3868	4	S47919	AF003325 Comamonas
36	56.6	3.4	7080	1	AF305325	AF305325 Streptomy
37	56.4	3.4	35654	1	SC7H1	AL021411 Thermotog
38	53.8	3.3	17641	1	AE001804	AE001804 Thermotog
39	53.2	3.2	1885	6	A48158	A48158 Sequence 5
40	51	3.1	134092	2	H0702G05	AL442105 Oryza sat
41	50.8	3.1	3663	10	RATSV2A	L01788 Rattus norv
42	50.8	3.1	3844	10	RATSV2AA	L05435 Rattus norv
43	50.2	3.1	2132	10	RNOCTK	X98334 R.norvegicu
44	49.8	3.0	12460	1	AE004903	AE004903 Pseudomon
45	49.2	3.0	3116	1	PPU10895	U10895 Pseudomonas

ALIGNMENTS

RESULT	1				
AF060173	AF060173	2421 bp	mrna	ROD	21-NOV-1998
LOCUS	Rattus norvegicus	SV2 related protein (SVOP) mRNA, complete cds.			
DEFINITION	Rattus norvegicus	SV2 related protein (SVOP) mRNA, complete cds.			
ACCESSION	AF060173				
VERSION	AF060173.1	GI:3901267			
KEYWORDS	Norway rat.				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 2421)				
AUTHORS	Janz,R., Hofmann,K. and Sudhof,T.C.				
TITLE	SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles				
JOURNAL	J. Neurosci. 18 (22), 9269-9281 (1998)				
MEDLINE	990197;5				
REFERENCE	2 (bases 1 to 2421)				
AUTHORS	Janz,R., Hofmann,K. and Sudhof,T.C.				

TITLE Direct Submission
JOURNAL Submitted (16-APR-1998) Center for Basic Neuroscience, Department of Molecular Genetics, HHMI, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA

FEATURES
source location/Qualifiers
1. .2421
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="brain"
1. .2421
/gene="SVOP"
301. .1947
/gene="SVOP"
/note="synaptic vesicle protein; similar to transport proteins"
/codon_start=1
/product="SV2 related protein"
/protein_id="AAC78627.1"
/db_xref="GI:3901268"
/translation="MEEDLFOLRQLPVVFRRTGESARSEIDDAASGEHDVQIEGRVRG
LEAVELDGAAPKPEFANPTDTEFVEDAVEAIGFRFQWKLISVLTGLAMMADAMEMH
ILSLAPQLHCHEMLPSQVALLTSVFVIGMSSSTLWGLNSIDYGRKTLGISVFWT
LYGILSAFAPVYSWILVRLGLVGFVGGVPOSVTLVAELFPMKARAKILLIEFWA
IGTFEVLAVFVPSLGRWLLLSAAPLLVFAVLCFWLPESARYDLVLSQNEKATA
TLKRIATENGAPMLGLKLIISKQEDRGKMRDLFTHPRWTFLLWFIWFSNAFSYGL
VLLTEGRFOAGDVCSISSRKAFAEAKSLACEYLSKEDYMDLLWTLSEFPQVLTLM
VIDRLGRKTMALCFVIFSLCSILLIFCIGRNVLTLLLFARAFISGGFOAYVYTFE
VYPATRALGLGTCSMARVGLITPFIAQWLESSVYLTAVYSGCLLIAALASCLF
PIETKGRALOBSSHREMGVGRGTNSVPRNSGSQE"

BASE COUNT 459 a 658 c 735 g 569 t

ORIGIN

Query Match 80.7%; Score 1326.2; DB 10; Length 2421;
Best Local Similarity 87.9%; Pred. No. 0;
Matches 1445; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 1 atggaggaggacttattccagtaaggcaagctgcgggtgtgaaattccgtgcacagc 60
DB 301 ATGGAGGAGGAGCTGTTCAGCTCAGGCAGTGTCCGGGTGGTGAATTCGGCGGCACAGGA 360
QY 61 gagagtgcgaagtcagagacacacgcttcagagagacatgaagtcacagattgaag 120
DB 361 GAGAGCGCAGCTCAGAGACACGCGGCTTCCGGGGAACATGATGTTCCAGATTGAGGG 420
QY 121 gtccagtgggcctgagagctgagagctgagatgagtgagggcagctgccccaaagagttl 180
DB 421 GTCCGAGTGGGCTAGAAAGCTCAGAGCTGGATGATGGAGCAGCTGTCCCCCAAGJAGTTT 480
QY 181 gccaatccccacagatgatactttcatggtggaagatgcagtggaagccattgcttggg 240
DB 481 GCCAATCCCACTGATGACACTTTCATGGTCAAGATGCGGTGGAGGCCATTGGGTTCGA 540
QY 241 aaatttcagtggaagctgctgtctctacatggcttggttgatg,jctaatgccatggag 300
DB 541 AGATTCCAGTGAAGCTCTCTCTCTCAGCGGCTTGGCTTGGATGGCGGACACGTCATGGAG 600
QY 301 atgatgatcctcagcatcctggcacacacagctgcatgagagtgagagctcccaagctg 360
DB 601 ATGATGATTTCTGAGCATCTCGCGGCTCAGCTGCATCTGGAGTGGCGGACTCCCGAGCTGG 660
QY 361 cagggtggcattgtagctcgtggtgtttgttaagcatgatgtccagatgccacgctctgg 420
DB 661 CAGGTGGCGCTGCTGACTTCGGTGGTGTCTTCATTTGGTATGATGTCCAGTTCTACGCTCTGG 720
QY 421 gaaatatctcagaccagctacgacgagaaacagggtgaagatcagctgctgagact 480
DB 721 GGAACATCTCGGATCAGTATGCGAGAAACAGGGCTGAAGATCAGTGTGTTCTGGACC 780
QY 481 ctgtactatggcatccttagtcttagtgcgccctgtagtgcgcatcctgtagctgctgctcgg 540
DB 781 CTGTACTACGGCATCTCAGCGCTTTCGGCGCCAGTGTATAGCTGGATCTCTGTGTCCTCA 840

QY 541 ggctctgtgggcttcggagatcggagaggttccccaagtcggtagcagctgtatgccaggttc 600
DB 841 GGCCTCGTGGGCTTTGGGATTTGGAGGGTGCCTCAGTCGGTGACCTGTACGCCGAGTTTC 900
QY 601 ctcccatgaaagccagagactaaatgtatttctgctgattgaggtattcttgggccaatcg 660
DB 901 CTCCCATGAAGGCCAGAGCCCAAGTGCATTTTCTGATTGAGGTTTCTGGGCCCATCGGG 960
QY 661 acagtgttcgaggtcgtcctcgtgctgtgttgatgcccagcctgggctgggcttggtg 720
DB 961 ACCGTGTTTCGAGGTGCTTCTGTGCTGTGTTGTATGCCAGCTGGGCTGGCGTGGCTG 1020
QY 721 ctcatcctctcagctgctccgctcctcctctcttcttgcgctgctgtgttcttgcctgaa 780
DB 1021 CTGCTGCTGCTCGGCCGCTCCACTACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1080
QY 781 agtcaaggtatgctgctcagggaaaccaggaagaaagcaatcgccaccttaaaagag 840
DB 1081 AGTGCTAGGTACGATGCTGCTGTCTGGGAACACAGAAAGGCCATTGCTTAAAGCGG 1140
QY 841 atagcaactgaaacgagctccccatgcgctggggaactcatcatctccagacagga 900
DB 1141 ATCGCAACAGAAATGGAGCCCATGCTCTCTGGGAAGCTCATCATCTCCAGACAGGAA 1200
QY 901 aaccagcaaaatgagagaccttttccacaccccatitttagatgacaaactttgctgctg 960
DB 1201 GACCGAGGCAAAATGAGGACCTTTTACACCCCACTTTCTGTTGGACAACCTCTGCTGCTG 1260
QY 961 tggtttatgatttcccaatgacttcttactacaggtattagtttactcaccacagaa 1020
DB 1261 TGGTTTATATGTTTCTCCAATGCCCTTCTTTATTACGGGCTGTCTGTCTGCTACCA 1320
QY 1021 ctcttcttgcaggagatgctgcggcactctccagtcgggaagagctgtagagcaaaa 1080
DB 1321 CTCTTCCAGCGCGAGATGTTTGCAGCATCTCCAGCCGGAAGAGCGGTGGAAGCAAA 1380
QY 1081 tgacgcttgccctgcagtaacctgagtgagagagattacatggaacttctgtgagacc 1140
DB 1381 TCGACGCTGGCTGTGAGTACCTCAGCAAGAGGATTACATGGACCTGCTGTGGACCACC 1440
QY 1141 ctctctgatttccaggtgctcctgtgactctgtggaattattgacgctggggcgcaag 1200
DB 1441 CTGTCTGAGTTCCAGGTGCTCTTGTGACTCTGTGGGTTCATCGACCGCTGGGCCCAAG 1500
QY 1201 aagacatgcccctgctgttctgtcactcttctcctcctcagcctcctgcttcttactgt 1260
DB 1501 AAGACCATGGCTGTGTTGCTGATCTTTCCTCTGTGAGCTCTCTGCTGTCTGTCTGTCTG 1560
QY 1261 gttggaagaaatgctcactcctgttactcttcttcttcttcttcttcttcttcttctt 1320
DB 1561 ATTTGGAAGAAATGTGCTAACCTCTTACTGTTTCTTTCATTTGCAAGAGCGTTTATTT 1620
QY 1321 ttccaagcggcctatgttttaacacactgaggtctacccacagcgaagcgccctcgagc 1380
DB 1621 TTCCAAGAGCGCTACGTTTTACAGCGCTGAGGTGTATCCAACGGCGACAGGGCGCTGG 1680
QY 1381 ctgggcacctgcagcgccatggcaagtggtgtgcttctcactcactccttctcagccag 1440
DB 1681 CTGGGCACCTGCAGCGCATGGGAGAGTGGCGGCTCATCACTCATTCATAGTCTCAG 1740
QY 1441 gtgagctggaatcctcctgtgtacctgactctggcaggtttacaggttgaggtgctgctcct 1500
DB 1741 GTGATGCTGGAATCTTCCGCTTACCTGACCTGGCCCTGTACAGTGGCTGTGCTGCTCT 1800
QY 1501 gctgcccctgctcctccttcttggccattgagacaaagcgagagctgagagagctcc 1560
DB 1801 GCTGCCCTTGGCTCTCTGCTTTCTGCCCATCGAGACCAAGGCCGAGCACTGCGAGGATCC 1660
QY 1561 agccaccgggagtgggccagagatggttcggccgaggaatgcacggtgcaggtgttaacc 1620
DB 1861 AGCCACGGGAGTGGGGCCAGGAGATGTTGGCCGAGGAGCAAAACAGACAGCGCTCCC 1920
QY 1621 aggtcgaaactctggctctcagga 1643

[illegible]

```
/protein_id="AAK65455.1"
/db_xref="GI:14523923"
/translation="MVKISNVRVRLVPLKOPYHWSYGIRESFAVNLIEIAIDJGV
GTGECTVAPDQTGTALILYRLAKHLVGHSPHDVAPLIARIFHOEYIIGHGANIMRAAQ
IPSGIDMAWMDLOGKLAGI.PVHOLLGGARHKAVGYFYFLOGETABELARDAAVGHAAQ
EFVFLYKVGREGKLELITAAVRGEIGDARLRLDANEGSWHDAIMCRKRLKYDIEF
IQPTVSNIPMAHVRKVGIPIVADQAAFTLYDYVEICRORADMICIGPREIGGI
QPMKAAVAEAAAGKICLHSSFTTGITTCAEHHIGLAIPNLDDGNOIMWQLVQEDIV
SSPDLTPKNGWLDAPFKPKGLFQALAEIDLVAEGREGRYAASR"
gene
CDS
complement(8311..9201)
/gene="Sma1462"
complement(8311..9201)
/gene="Sma1462"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; global similarity to
choline-binding protein precursor from B. subtilis
(GI:7404430) and several putative ABC transporter
proteins; adjacent ORFs encode other putative members of
the choline transport system"
Query Match 5.5%; Score 91; DB 1; Length 11856;
Best Local Similarity 48.3%; Pred. No. 7.4e-13;
Matches 324; Conservative 0; Mismatches 335; Indels 12; Gaps 2;
Qy 199 acttctcatggtgaagatgcagtggaagccattggctttggaaaaatttcagtggaagctg 258
Db 1423 ACAATCACACTAGACGATGCGCTCGATCGCGCGCGACGGGAACCTTACCAGCGCAGACTG 1364
Qy 259 tctgtctcactggtggttggttgatgctgagatgcacatgagatgatgattcagtcac 318
Db 1363 ATGGCCATATTCGTGTTGGCTGGCGATGCCATGCCATGAGGTGCTGGCGGTGGCTTC 1304
Qy 319 ctggaccacacagctgcatctgcagtgagagctcccaagctggcagtggtgctgctgacc 378
Db 1303 ACCGCCCATGATCGCGCGCCACTTTCGGCCTGCACAGTTCGCCGAGGCACCTTCAGACGGCG 1244
Qy 379 tcgggtgctttgtaggcatgatgcacgtccacgctctcgggaaatactcagaccag 438
Db 1243 ACCGTTTCTTCTCGGAATGCTGTTCGGCGCGCGCGGTTTCGCGAGGCTCGCGGACCGG 1184
Qy 439 tacggaggaacacagggctgaagatcagcgtgctggtgagctctgtactatggtcattcct 498
Db 1183 ATCGCGCGCGCGCGCTTCGTATCGCCACCGTCGCTCGCATGCCGTCCTTCGGTCTGCTG 1124
Qy 499 agtgatttgcgcctgtatagctgatactcgtggtgctccggggcctggtggcttcggg 558
Db 1123 TCCGCTCTTTCGCGAGGACTTACCGTTTTCGCTGCTGCGGTTTCTCAGCGCGCGAGCT 1064
Qy 559 atcggaag---agttcccccagtcggtgacgctgtatccgagttcctcccatgaaagcc 615
Db 1063 GTCGCGCGCACATGCCCGTGCAGTATGCGATGATGCGGANGTCTTCGCGCGCGCGAAC 1004
Qy 616 agagctaaatgatttgcgtgattgaggttattctgggccatcggagcagtggttcaggtc 675
Db 1003 CGCGCGCGCTGGCTGGTGCATGCTGGAAGGCTTCTGGGCTGTGCGGACGCTCATCGTGGCG 944
Qy 676 gtctgactgttctgctgagtcaccagc-----ctgggctgagcttggtgctgctc 726
Db 943 CTCGCCGCTTGGGTGCGAGCCTCGCGCGGCTCGCGGATGCTGCGGATGATCATCTTCGCA 884
Qy 727 ctctcagctgtcccgcctctctctcttggcgtgctggttcttctggtgctgaaagtga 786
Db 883 GTGACGCCAATCCCGGCACATGATCGCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824
Qy 787 aggtatgattgtctgcagggaaccagaaaggcaatcgccaccttaagagagatagca 846
Db 823 CTTTATCTGTCGCCCTGGGAAAGACCTCCGAGGCGAAGGCCATCTCGTATGATGATCCTC 764
Qy 847 actgaaacagg 857
Db 763 GTCGTCAACGG 753
```

```
RESULT 5
AC004642 148432 bp DNA sequence (pls DS00543 (D193) and
LOCUS Drosophila melanogaster DNA sequence (pls DS00543 (D193) and
DEFINITION DS02867 (D200)), complete sequence.
ACCESSION AC004642 AC003551 AC003552 AC003553 AC003554 AC003555
AC003556 AC003557 AC003558 AC003559 AC003560 AC003561 AC003562
AC003563 AC003564 AC003565 AC003566 AC003567 AC003568
AC003569 AC003570 AC003571 AC003572 AC003573 AC003574
AC004642.1 GI:3097827
VERSION HTG.
KEYWORDS Drosophila melanogaster (Subclones in Sac from Pl clones DS00543
SOURCE (D193) and DS02867 (D200)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 148432)
AUTHORS Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Humastti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,D., Snir,E.,
Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
and Kimmel,B.
TITLE Sequencing of Drosophila chromosome 2R, region 60B1-60B10
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 148432)
AUTHORS Celniker,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Humastti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,D., Snir,E.,
Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
and Kimmel,B.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS
COMMENT 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site
(http://fruitfly.berkeley.edu/sequence-archive.html) or send email
to drosophila@hgc.lbl.gov.
Library locations: 63.6, 83.30.
FEATURES
source
1. 148432
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="60B1-60B10"
/clone="pls DS00543 (D193) and DS02867 (D200)"
/notes="d193 extends from bp 1 to bp 83271 and d200 extends
from bp 81802 to bp 148432."
BASE COUNT 38713 a 34662 c 35252 g 39805 t
ORIGIN
Query Match 5.4%; Score 89.4; DB 3; Length 148432;
Best Local Similarity 59.1%; Pred. No. 1.8e-12;
Matches 153; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
Qy 196 gatactttcatggttggaagatgcagtggaagccattggttggaatttcagtggaag 255
Db 75185 GATACCTACACTGTTTCAGCAGCGCATCAATGCCCTTCGGATGCTTTCAGTCAAG 75244
Qy 256 ctgctgttctcactggttggttggttggttggttggttggttggttggttggttggt 315
```

Db	75245	CTGTCGCTCTTGGTGGGCTCTGCTGATGTCGAGCTCCAGTAAATCGCCATTTCTCAGC	75304
QY	316	atcctggaccacagctgcattgcagtgagagctcccaagctggcaggctgctcgtg	375
Db	75305	ATATTGGCCCGCTTGTTCGCGAGTGAATGTCACCAAGTTCACAGCAGCATCCGTC	75364
QY	376	acctcggctgtctttagcattgatgtccagctccacgctctctgggaaatatctcagac	435
Db	75365	ACAACAGTGTGTTTGGCGATGATGTTGAGTTCCTCAGTTTTTGACACAGCTGAGCAAC	75424
QY	436	cagtcaggcaggaaaacag	454
Db	75425	AGATATGGCGGAAGTCGG	75443
RESULT	6		
AC020509	AC020509	157851 bp	HTG 03-JAN-2000
LOCUS	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***		in ordered pieces.
DEFINITION			
ACCESSION	AC020509	1	GI:6664388
VERSION	AC020509.1		
KEYWORDS	HTG; HTGS_PHASE2.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
REFERENCE			
AUTHORS	Adams, M. and Venter, J.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA		
COMMENT	This sequence was identified as CUM:10213488 by the submitter. For more information on this record e-mail to fly@celera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.		
FEATURES	Location/Qualifiers		
source	1. 157851		
BASE COUNT	41166 a 36910 c 37402 g 42373 t		
ORIGIN			
Query Match	5.48; Score 89.4; DB 2; Length 157851;		
Best Local Similarity	59.18; Pred. No. 1.8e-12;		
Matches 153; Conservative	0; Mismatches 106; Indels 0; Gaps 0;		
QY	196	gatactttcattggtggaagatgcagtggaagccattgctttggaaaatttcagtggaag	255
Db	107182	GATACCTACACTGTCACAGCGCATCAATCCCTTCGGTTCGGAAGTTTCACGTC	107241
QY	256	ctgtctgtctcactggttggcttgatggtgattgctgcatgagatgatgattcctcagc	315
Db	107242	CTGTGCGCTCTTGGTGGGCTCTGCTGGATGTCGCACTCCATGGAATGGCCATTCTCAGC	107301
QY	316	atcctggaccacagctcattgcagtgagggtctcccaagctggcaggtggcattgctg	375
Db	107302	ATATTGGCCCGCTTGTTCGCGAGTGAATGTCACCAAGTTCACAGCAGCATCCGTC	107361
QY	376	acctcgggtggtctttagcattgatgtccagctccacgctctgggaaatatctcagac	435
Db	107362	ACAACAGTGTGTTTGGCGATGATGTTGAGTTCCTCAGTTTTTGACACACAGCTGAGCAAC	107421
QY	436	cagtcaggcaggaaaacag	454
Db	107422	AGATATGGCGGAAGTCGG	107440
RESULT	7		

AE003462/c
LOCUS
DEFINITION
of 15, complete sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
mRNA
gene

AE003462 300542 bp DNA INV 04-OCT-2000
Drosophila melanogaster genomic scaffold 142000013386038 section 11
of 15, complete sequence.
AE003462 AE002575
AE003462.1 GI:7291637
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 300542)
Adams, M.D., Celnik, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amaratunga, P.G., Scher, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor,
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
Andrews-Pfannkuch, C., Balgwin, D., Balles, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkov, D.,
Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
Fleischmann, M., Fosler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Krivitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Palazzolo, M., Pittman, G.S., Pan, S.,
Nusskern, D.R., Pacle, J.M., Palazzolo, M., Reinert, K., Remington, K.,
Pollard, J., Puri, V., Reese, M.G., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Saunders, R.D., Scheeler, F., Shen, H., Spier, E., Spradling, A.C.,
Simpson, M., Skupski, M.P., Smith, T., Svirskas, R., Tector, C., Turner, R.,
Stapleton, M., Strong, R., Sun, E., Svirskas, R., Wasserman, D.A.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Williams, S.M., Woodage, T.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Ye, R.F.,
Zaveri, J.S., Zhang, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster
Science 287 (3461), 2185-2195 (2000)
20196006
2 (bases 1 to 300542)
Adams, M.D., Celnik, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
Location/Qualifiers
1. 300542
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
complement(join(8417..8923,14564..14687,19965..>20155))
/gene="CG11169"
/product="CT31208"
/db_xref="FLYBASE:FBan0011169"
/db_xref="FLYBASE:FBgn0034905"
complement(<8417..>20155)
/gene="CG11169"
/map="60A1-60A2"
/db_xref="FLYBASE:FBan0011169"

Db	238959	ACAACAGTGGTCTTTTGGCATGATGTTGAGTTCAGTTTGGACACAGCTGAGCAAC	238900
QY	436	cagIacgacgaqaacac 454	
Db	238899	AGATATGGCCGGAAGTCGG	238881
RESULT	8		
AE001863			
LOCUS	AE001863	198677 bp	DNA
DEFINITION	Deinococcus radiodurans R1 section 2 of 2 of the complete chromosome 2.	BCT	22-NOV-1999
ACCESSION	AE001863	AE001863	
VERSION	AE001863.1	GI:6460670	
KEYWORDS			
SOURCE	Deinococcus radiodurans.		
ORGANISM	Deinococcus radiodurans		
REFERENCE	Bacteria: Thermus/Deinococcus group; Deinococcales; Deinococcus.		
AUTHORS	1 (bases 1 to 198677) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.		
TITLE	Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1		
JOURNAL	Science 286 (5444), 1571-1577 (1999)		
MEDLINE	20036896		
REFERENCE	2 (bases 1 to 198677) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA		
FEATURES	Location/Qualifiers		
source	1..198677 /organism="Deinococcus radiodurans" /strain="R1" /db_xref="taxon:1299" /chromosome="2" complement(91..1926) /gene="DRA0206" complement(91..1926) /note="similar to GR:AE000520 percent identity: 43.92; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="oligoendopeptidase F, putative" /protein_id="AAF12423.1" /db_xref="GI:6460718"		
gene	/translation="MLRACPLRGAGRSRCPAPSQTCSPOCMERHWRALPPMTTQ SNLEAVEHLKVPADDAQWASVAFALQADLTAGDVPAWLERWNSLSARLQSGS KLSTHADLTDDADTQORYQTFAEVLPAERASOALTEKLAVPCYVPGDFALNRY RFQDAALFRANVDLGYTHNQIINRHSVITGNKVLGOELTLPQAKQOMDSPDRA OREAWRALSNAAIAPELDAMLDLIRTRQLAHNADEANFRYRKELDRDYTP EQCLDFASVDEVPVLLSLGVDIARGLLESRPNDYNNRNLLDPGERSLRFPAF GADLALAQQYTESLDPLAARFQMRRESGLDLESRPGKMTATCNFYPTNEPFL MNVGAEADRLVFEHVEGHGFYSGDRQLVNNRWSPIEFVEIPSNAMAEFLTDHL SHVFSDDLELARYRRKQLEGVIAFLPWAQMDAFQHWLYAEAPEDVTIADLAKWLELD RTYHPFDWSGLDESVRAGKWHYHIFQVPFYIEYAMCYLAALAAVIRAAQRDPAAL GRYKDALRLGNTVSPELYRAAGAEFEREDREYIRGLMGFLKEQLG" complement(2005..3645) /gene="DRA0207" complement(2005..3645) /note="similar to PID:1653945 percent identity: 49.28;		
gene	identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12422.1" /db_xref="GI:6460717" /translation="NNFEVLSPSLSSVAHALNRASLKPEMMSKRSLSALLGALLL SACGSPENGQAPAPOLPNPTEEAEGTSGKVRGLWDAFGPGIKTPEEVSTLVDTAR AMNNTLVQVRRGDCYCNNAAMPRTDDPEVPAGFDPLADVIKHAHAKDIEVNAWII TALMNSVTPPKSAEHFNTGPNAGARDNNMLTVKADGTVKAGADWLDPGHPDAAAY IRNMYVSVAKNYAVDNIIMFDRVYDPYNPVGQVQMGYNETALARYRAETGATGDPD ADPQSAWRROQVNTLRVETALAVKAPGVAIAATITTYGEGDAEAFASRYPYAE VGDKLWTKSYGLDQNMNMYKRDYFVPLQAGWQDQWNTFAAGLRORYPHTTQVSGAA IYLNDIASVSOIRKTYQAAGLWAGYSYETPKDYNECKRGAEAELADLKLRLTDAGG PFTFVKHAPLPAASVRGLSGELRAARPGHSHVELYVRAGCTWNDVVAATFSDGGRYG FLNLPQAGFDRVVDGKILLPNVQATPGRVTMLPEITLN" complement(3751..4860) /gene="DRA0208" complement(3751..4860) /note="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="peptide ABC transporter, permease protein" /protein_id="AAF12421.1" /db_xref="GI:6460716" /translation="MTGGPETPTPLQMAWRRYKRKSPGVIAGWLIALLVIALSPFL APYNTTQHEGYEQRPQTHLHQRRLHRPEYGVSTTKRDFQVTAFTAKTVPDTHPLP VOFFYKGPFSYGLQSDRHLFGVDGDEYGEKVPYFFLEGTDLGRDLFSRTLVG AQVMTVGVVGLISFALGILGVSQYFGWVDLQIRVVELVLSFPKPLTILALAT LIPAMPSTWYTGIVAVLALIGWASLARVRGQVLAARNLDYVSAQAAMGASNLRI LPHITPNLSSFLVTATLALPGYILGESTLFLGIGLKEPMTSWGLLLRUASKLEVVS SYPLLWGLFVFSVLAFFNFGDALRDADTQSR" complement(4857..5843) /gene="DRA0209" complement(4857..5843) /note="similar to PID:1001492 PID:1001545 percent identity: 62.93; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="peptide ABC transporter, permease protein" /protein_id="AAF12420.1" /db_xref="GI:6460715" /translation="MLCCIMLNVTLRILGMPTLLVSVLCFLVKIOPGSGFIDQYL EDPRSKETVANITRQLGLDQPAWKQYLNWITGVVTRLDGYSFASAPVTSIGERL GMTVFIATLTALTWLIAVPLGITYAFNRYGPAQIANFLGVGLAIPFLAALLLV LVRLQGGTVNGGLFRPDMIDAPWFAKVIDLLAHLWIPVLAVGLEGVAGLMQMRAS LDVLSQDYIRTAKAGLPGQVRVWKHVRNAINPLISLAGLSLPTLISGTTIISVMS LPTIGLLYDSLLNKDQYATMTLLMLSAFLLILGNLLSLDALAWADPRVRS" complement(5964..7804) /gene="DRA0210" /note="this region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AB000657 percent identity: 51.19; identified by sequence similarity; putative" 7764..8603 /gene="DRA0211" 7764..8603 /note="similar to GB:AL009126 percent identity: 60.78; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transcriptional regulator, GntR family" /protein_id="AAF12419.1" /db_xref="GI:6460714" /translation="MWSNNKTFSELVRKWKYGCVSSPAPLPPEPPANWVALDPADAP AYVQLAQLRRRIEESGASLPAERELARGLVSRVTLRQGLLEAGELLRRRR GSGTFVAGTPHTSRQLGLLSFSDVEVRAQGVPGARVLSFTQGRTPQALSLGLTAA		

YVLPYFSTYAAASAASOETLMGFCAGYGVFLGFCYNYLFLFVLSREATYAWYL.YNL
SIGLFSASDGLFLKLLPHGVAVESAGIYGVFLGFCYNYLFLFVLSREATYAWYL.YNL
FRUGLLACVLLTASGLPLVGLRAMVNLASVLMVLSLLSLLAGVHVRGOLRYGLYI
LAWGALLLSFLVTTASGLFELFGVSSVYKIGMTVELTSLIGLADINAKKEGF
RSQAAEQARVENEAKRSRLAKMSHEIRTPNLVGLMLQLLQRLDTPDRGQAAVETIA
SSGSAVINDILOYARIESCKHLERIDPLELLSOTLALFSAQAVEKRLHLHLG
LQGVPRRLNGDPTRLRLQVMLNLSNAKLFYAECHVAVPVRREFDEGGERLLYSVSD
SGIGTSAQAKTLFESQASDSTTRRYGSGGLAISRELQVMGGRLEYSSEBKG
TRFSDVLPALDADEADELAQIOLPALLASEJNUTDCLQALLERKMGURVERCL
QBRLLNAYLEDSAPLLVLAAPWFGPPSILWDTLYAQEQORILLCPPOHCOQLP
PIEGLRLAQAQVAVKALREALLERYRRRQPGRSSEAPDRPPDAPCILVAEDN
PVNOLVRCFLAKRGVAVRAGNLGRALIDEYLRDPNGIOLILMDGEMPMQGFATRL
IREPRAQGWPRPPIVALTAHILDEHRRAGIEAGMDAYLGKPKVDRAELVATLERLLQG
PSRQA"
2911..3141
/gene="PA3463"
2911..3141
/gene="PA3463"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG06851.1"
/db_xref="GI:9949606"
/translation="MLIPHDLEADTLNNLLEDFVTRECTDNGDETPLDVRVERARHA
LRGEAVILFDPESQOCQLMLRSEVPAELLRD"
complement(3959..5284)
/gene="PA3464"
complement(3959..5284)
/gene="PA3464"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG06852.1"
/db_xref="GI:9949607"
/translation="MKLLRCLLLGAVVSPSLWANSNHTVGSYLALRELAALREAPE
VEVEPLFAALAEERGLAULIDQEAARAHIGNYPARPDALRFAEGERGDLKQAF
AALRNPEITLALALQPLQDQORPHKPOEVLFVFNLSPTWAFIRKLEPGERVA
PLAVATADEPDYGHIDINLFDONPGAGQYGFQPGDPRFESSQAPFHMGYH
EAVIYSGAPFLARWEMRAYOYFGLSFANFANGHYWYRFLGWMHYIODITOPY
HSTPLCASALOMEGKALLGYPEEKAAIERVANRHTAVEKYOFDMIRQLRDRGR
POMLDAADTRDQGPAYSPYLRDVVAASNAHAADFDAIGWLAARPASAAQD
FSSNQPRPAHDNAVNLNAQLIELIGHFGAHSNRNARALETEEGAKE"
5501..7216
/gene="PA3465"
5501..7216
/gene="PA3465"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG06853.1"
/db_xref="GI:9949608"
/translation="MQGSVPKHDPLKPEGQAAPERTGTWAPFSTQAPRIIWIWML
FANLGTWQSVAAWVYTDHASPVMAMTQVAAALPLVLLSLSGVIAADNHDHRRKIM
LWGLSFEMTAMFATLLAFGLYLDPVLLIISILWISLGSVTIPAWQAIVNEQVPRM
VSDAVLLSNVYNVARAAGPALGLLSAVGPAWVFLNFSFCYMLIAWIAWQWRDVP
KRSLPPEGLIGVTAARFTQYSTVTRLVNMRSAFGLSASAVWALLPLLAHNPQD
AIIYGMALGALGATLGSVSRLEORIGSRSLISAGFTLALILLTLGLVNDLWL
FVLLIGCGWIGALATYNSAVQILYVDPNWKARALAYQYALYGGALGSLFGLHAE
TMVHGALLAAGCULLLASVILLNSRLPMDAASISRAPASMPQPSFVFNTRRGWYL
VSIERYIPAERTDRFRAAQLRLRLRNGAERWSLPRDVSNPPEVWQELFLVDNIQH
LRMLDRMTLADKIVINDVNTALHAGDGPPIRHCVSYEASSYDTPLVKASATPPANDEAG
ATAGS"
complement(7338..8678)
/gene="PA3466"
complement(7338..8678)
/gene="PA3466"
/codon_start=1
/transl_table=11
/product="probable ATP-dependent RNA helicase"
/protein_id="AAG06854.1"
/db_xref="GI:9949609"
/translation="MPSRPLFSQFPLHERLLKALELSFSEPTPVQAAAIKALEGHD
LRVTAQTGSGKTAALFLPLHRLLAEDKPKSLARALLPLTPRELAAQQTLEVERFAQF

TFIKACLITGGEDFKVQGARLKRKNPEIITGTPRLLLEORNAGNLPLODI EVLVLDEAD
RMLDGFADVDLALANACPAEROTLLFSATHSGAGLNKVI AEVLREPOVRLRNOGEL
NENPQOVIITADDAHAKQLOMLLSNETYTKAIVFTNTRASADRLGRLIANQHVF
VLHGKKDKRKLAIERLQGNVKTIVATDVAARGLDVEGLDLVNFDMPSGSDYVH
RIGRTGRAGAEGLAISLICHGDMNLMSSVERLYKONFERNNIKELKAAQYQPKLKAS
GKAAGSKKKQDRKGAAPAAKRAKPAARPAKPSAVSVADGMAPLKRKKPTAE"
8864..10231
/gene="PA3467"
8864..10231
/gene="PA3467"
/codon_start=1
/transl_table=11
/product="probable MFS transporter"
/protein_id="AAG06855.1"
/db_xref="GI:9949610"
/translation="MPDAPASAAERLERLPLSPYHRLVFVILALAFFDMSMDLAMTF
LLGSKIAEFGLDLSAAGLLASSFFGMVIGALSGMLADRFGRKVPFOASIVLWGLAS
YLCTAGDLDLSITFYRLIGIGMGEFFPIAQSLSEMIPASRRGXIALMDGFWPLGF
VAAAGLSYFLPLTGMRSIFLVAIPAVFVLAIRFLIPESRMLQAGRQADRVLR
DIEARVMRSGLTELPPURQPORERSRPFSAELMSPAYRRRTITVWGLWLFAL
LGFGYGLTSMLSALLQSGFPAVTSVYTVLISLAGIPGFLCAAMLVESWGRKPCVLM
LLGGAMAYAYGOTAVFGGSLALLIGFLAMOFFLFGMMAVLYTTPELYPTTSARATG
SGFASAVGRIGSLGLPLVTGLVPLTGGGVFTLGALCFGVAALVVMARFIEIETGRTL
EELAG"
BASE COUNT 1495 a 3698 c 3395 g 1756 t
ORIGIN
Query Match 5.0%; Score 82.8; DB 1; Length 10344;
Best Local Similarity 47.6%; Pred. No. le-10;
Matches 277; Conservative 0; Mismatches 302; Indels 3; Gaps 1;
QY 243 atttcagtgaaagctgtctgtctcactgctgtggttgatggtgatgccatggaagat 302
Db 8920 ATATCACCGACTGTCTTCGTGATCATCGCGCTGCTTCTTTCAGATCGTGACCT 8979
QY 303 gatgatcctcagatcctcgtgaccacacacagcattgagcagtgagcgtcccaagctgca 362
Db 8980 GGCAGATGACCTTCTGCTCGGCTCGATCAACACCGGATTCGGGCTCGACCTCGCGCA 9039
QY 363 ggtagcattgtgacctcgggtggttctttagcattgatgctccagctccagctctgggg 422
Db 9040 GCGCGGCTCTGCGCAGCTCGAGCTTCTTCGGCATGTGATCGCGCGGCTGTCGGG 9099
QY 423 aaatatctcagcagcagtcacgacgagaaacagggctgaagatcagcgtgctggaactct 482
Db 9100 CATGCTCGCCGACCGCTTCGGGCGCAAGCCGGTATTCAGGCCAGCATCGTGTCTGGG 9159
QY 483 gtactatggcatccttagtgcatcttgcccgctgtatagctggatccttggtctccgggg 542
Db 9160 CCTGGCAGCTACCTGTCTCCACCGCGCGGACCTTCGACAGCCTGACCTTCTACCGGT 9219
QY 543 cctggtggttcctgggacg ---aggagttccccagtcggtgacgctgtatgcagagt 599
Db 9220 ATTGCTGGGATCGGCATGGGATGGAGTTTCCCATCCCGACGCTGCTGTCTCGGAGAT 9279
QY 600 cctcccatgaaagccagagctaaatgtatttctgattgaggtattctgggacatcgg 659
Db 9280 GATCCCGCCACCGCGCGCAAGTACATTGCGCTGATGAGCAGCGCTTCTGGCGCTGGG 9339
QY 660 gacagtgctogaggtcgtcctcctggtgtgttctgtagtccccagcctgggctggtgct 719
Db 9340 CTTGCTGCTGCGGCTGCTCTACTTCTGCTGCGCGTGCACCGCTGCGCGCAGCAT 9399
QY 720 gctcactctcagctcctccctctcttgcgctgctgtgttcttctgctgctgctga 779
Db 9400 TTTCTGCTGCTGCGGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9459
QY 780 aagtgcaggttatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 821
Db 9460 GTCG :CGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9501

RESULT 11

```

AC067757/c
LOCUS       AC067757       62204 bp      DNA          HTC          27-APR-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-423G3 map 2, LOW-PASS SEQUENCE
SAMPLING.
AC067757
AC067757.1  GI:7651806
KEYWORDS    HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 62204)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karalas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 19993
Center clone name: 423_G_3
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 730: contig of 730 bp in length
* 731 830: gap of 100 bp
* 831 1522: contig of 692 bp in length
* 1523 1622: gap of 100 bp
* 1623 2322: contig of 700 bp in length
* 2323 2422: gap of 100 bp
* 2423 3119: contig of 697 bp in length
* 3120 3219: gap of 100 bp
* 3220 3925: contig of 706 bp in length

```



```

KVAAAITDATTNADAMSVTLRLDTAAASLTVNAQDSQGRILANGTASFPAGAAET
TATVAPFELNDRFARUSIEGVSAGAVHLDDGFRRRRVALLAGETGNDQFPLLOPL
YYISRALQFVDLIPQROADLAQAIPEILOQNSPVIVMAQIGRLPOETIYPLQRWLAG
GOTLIRFAGPKLAAPADDPVPLVTLROGERALGALSWAEPPLADPAFAGAGMP
KADGLVLRQVLAERTIPDLAERTWASLADGTPLVTRNVEAGRIVLHVSAEASMSDL
PLSCHEVDMLARIIVOLSKAGASANAPATALPPFRLTLAEGAMSAEIGTARPLEMR
AQOPRTGDFNPGLGYCTGDEGEVALNLLPAGATRLPDLTSAAGVSTTNEALIGETAKS
LRPALPTAAMMLIADSLIVLEFMGCAFARLPKARSATVATLLALGAFAMSPDARA
DDPKGDSQIIFERLDTTHLAYVLEGGEDVDRISEOGLOGSEFLTWRTITLBPQPVGL
DISKDELSFYPIIYWPVSATAMPSSAAISRIDATYMRAGGTIVLFDTRQOFLSSOIGAT
SANGERLQAILANIDIPPELVPEHDVILTRSFYLTNFPGRYNGSPLWVESRQGAAT
TSLGSSGDGVTPMITGNDFAGAMVDAQAPVLTVPDEMORHEAFRSGVNTIMY
MLTGNKADQVHVPALLERLQ"
4888.. .6972
/gene="Agr_C_4105"
4888.. .6972
/gene="Agr_C_4105"
/gene="Agr_C_4105"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_4105p"
/protein_id="AAK87998.1"
/db_xref="GI:15157412"
/translacion="MTLFAPELPMIVIAILAVAALSLF IGLWRGVGKAMLRGLAIY
ALLAAIANPLLTQEEREPSTLIPVIVDRSQSDQDRPQMTDTALETLKDRLSRPPR
IEPRIVEVDNGESDPSQTLFSAISSAVSDVPSRVGGAIFLSDQIHDIPNALPNA
EQALGPRAPVHGLITGKADEPDRRIEILVPAPRFGIVNFEQOLTLRVDFDGRPAGGSA
EVTVMNGEELATLQATPCQPTFFSKVPRGNNWMEFSAELPELGEVTAANNRAVHLI
EGTRONRLVYSGEAGERAWNNLLKSDASVDLVHFTILRPPRQDGTPINELSLI
APPTRELFEKINDEFLIIFDRYQHRGVLPIIYDYIAQYVEKG ALLIAAGPEHAGQ
FRTDPLASVPLPAQGTQVHQAAFPRLSEAGKHVPTRGLESGNPEPPGRMFRFT
ISVDPQGTQVLMGDQVPLLVNROGEGRVAMLLSDQGLMARGFEGGSGPHVALYRR
IAHMLKEPELEALTARANGRTLIQITROTIGDQGSVRFPSCKTENMAFTTTEP
GLYRIERMDTEGLFKNGDRTLLVHVCADVADPEKAMISTETLKLPLADATKGSVH
LRVADNSGRIALDPLVRGDIRVADEDRMLIRLTDTVLKGVNTLPLPAGFAGLALL
LAFSALWREGR"
7017.. .7475
/gene="AGR_C_4106"
7017.. .7475
/gene="AGR_C_4106"
/note="(L25828) ORF 12"
/codon_start=1
/transl_table=11
/product="AGR_C_4106p"
/protein_id="AAK87999.1"
/db_xref="GI:15157413"
/translacion="MRPKGCAVRRIEKNMTNFVLSDVADAEAKAIRDPLVAYNLARF
JESDKRDLNITIRNDNSVTGGIVGHTARGWLYVOLLFVPEAMKGGIAPKLLAMAEF
EARKRCMGAYITDMPDALTARYERYGFTKIGSLGPLSSQSGITWLEKRF"
complement(7485.. .8798)
/gene="AGR_C_4107"
complement(7485.. .8798)
/gene="AGR_C_4107"
/note="probable sugar transporter - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_4107p"
/protein_id="AAK88000.1"
/db_xref="GI:15157414"
/translacion="MTAITNDALERAGAGVORRLMIGFLVWAADAMOVLAQVTA
ASTAATGTLVPAALGTAFVFCMLLGAAGFCGLADRYGRRRLVITVACDALFGVL
STFSPDTILLVRLFLIGAVAGGTLPLVDYAMMAEFPAKNRGRHVLFECPVAVGTLI
VALAAGVAGVADAWRYIFAVTAFPAVLGLRLVLPESLYLLRTGRSEAKAVV
NRMLVNGRAPLDAGTGLFOPDTKGGIFSPALRQSRMMLIAIWLFSVSYVGVFTW
MPAKLAGDGFVRGFLVLVALAQIPGVAIAAYGVKGRKPTLIGFCLLSALGCL
LFTLASAGAMIASLLIMSFPALLTGWALYAFYPTESRATGMGAAMARLGLL
LAPSLGVAIAQCGFVAGIGIFAGLLVLAALATMINAETROVALT"
complement(8888.. .9493)
/gene="AGR_C_4108"
complement(8888.. .9493)
/gene="AGR_C_4108"
/note="probable acetyltransferase VC0655 {imported}"

```

```

Vibrio cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_C_4108p"
/protein_id="AAK88001.1"
/db_xref="GI:15157415"
/translacion="MSPHSGSGFLFMSKHDLVYLTEDASHDAIIDLINEAFQGRH
TRAAARIRFOGPHDRSISFCVADDDGETIASVRMTPLVLAGAVKHMGLGPLAVRPSHKNM
GIGRELVRIAIAAAKRGKSEAVILIGPPPYMPLGFEKVAYAAALDFPQVDPNRLVLV
PVGDVHORLKGKIGWRKCEAAMPAAAEDESEFEPLRVYG"
9586.. .10746
/gene="AGR_C_4109"
9586.. .10746
/gene="AGR_C_4109"
/note="HYPOTHEICAL 37.4 KD PROTEIN IN EXUR-TDCC
INTERGENIC REGION (O328)"
/codon_start=1
/transl_table=11
/product="AGR_C_4109p"
/protein_id="AAK88002.1"
/db_xref="GI:15157416"
/translacion="MKIAIIRKIIPRQIISRPSHSARAGPETHCSMPGPAASRKAG
RIAVKETIKEIVMGLVGVVKMDVMDYDTKETKGHKRSASQFRNMVTVDGAVPSGEG
GFEAERKDRYHLVSVFACVPAHRTLIFRKLKLEDLISVSVVDPLMLENGWFKPGEGER
TPGATQDHLFGSSSTLQWYTKADPTYSGRVTVPVLDKDKRCTIVSNESAEIIRVFNVA
FNDLTGSADADYYPKALRAEIDALNDVIYDVTNNGVYRAGFATTQDAYQONNAKVPETL
DMLEDRLKDRRFLFEGSEQTEADWRLFTTLVRFDPVYVGHFKCNIRIHDYPHLTGTLR
DLYOTPGVPDVTVDMRHIKEHYVRSRHTINPTGIPIGPELDTLMAHGREALS"
complement(10764.. .11192)
/gene="AGR_C_4111"
complement(10764.. .11192)
/gene="AGR_C_4111"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_4111p"
/protein_id="AAK88003.1"
/db_xref="GI:15157417"
/translacion="MHFVFLIIRGATLGVRAACFDEKGRIFLVRHLYPLGMYLPGGGV
ERGETLLMALHKEIRRECNLDAAISPELVHVLNLEGSNRDHRVALYRLOVOTTTPKKP
DHEITESGFFELSALEPEGLTPATRRRLAEAGEAAIADYW"
BASE COUNT 2250 a 3769 c 3294 g 1973 t
ORIGIN

Query Match 4.7%; Score 77; DB 1; Length 11286;
Best Local Similarity 48.0%; Pred. No. 3.3e-09;
Matches 320; Conservative 0; Mismatches 335; Indels 12; Gaps 3;

Qy 203 tcatgttggaagatgcagttgaaagccatggcttltggaaaaatttcagtggaagctgtctg 262
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8788 TCACCATGAACGATGCGCTGGAACGCGCGGTGCGCGGCTATCAACGCGGCTGATGG 8729
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 263 ttctcactggcttggttgatgagtgatgccatggagatgatgactcagatcctctg 322
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8728 GCATATTCGCGCTCGTCTGGCGCGCGATGCCATGCAGGTTCTCGCGGTCTTCACAG 8669
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 323 caccacagctgcattgcgagtgagagctcccaagctgcagagtgagcatgctgacctcgg 382
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8668 CCGCCCTCATGTGCGCGGACCTTTGGCCTGACAGTGCAGCGCAGGCGCTGCACACCGGACGG 8609
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 383 tggctctttagcatgatgtccagctccagctccagctctctggggaataatctcagaccagta 442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8608 CGTCTCTTTTCGGCATGTTCTGGGTGCTCGCGGTTCGGCGGTTCGGCGGCTATG 8549
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 443 gcaggaaacaggctgaagatcagctgctgtggaactctgtactatggactccttagtg 502
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8548 GGCGCCCGCGGTGCTGATGCTCACCCTGCGCATGCGCTGTTCGGCGGTCTTCGA 8489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 503 catttcgcccgtgtatagctggatcctggtgctccggggcgctgggtggctcgggacgc 562
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8488 TATTTTCACCGGATTTTACCATTCTGCTGCTCCGCGCTTTTCTCACCGGAGCTGCGGTGG 8429
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


CC The orientation of the sequence is from SP6 to T7 end of the P1
CC clone.
CC Genes were identified by a combination of three methods: Gene
CC prediction programs including GRAIL (available by anonymous ftp
CC from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
CC Washington), Genscan (Chris Burge,
CC <http://quomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene
CC (<http://www.cbs.dtu.dk/netgene/cbsnetpgene.html>), searches of the
CC complete sequence against a peptide database and the Arabidopsis
CC EST database at TIGR (<http://www.tigr.org/tldb/at/at.html>).
CC Annotated genes are named to indicate the level of evidence for
CC their annotation. Genes with similarity to other proteins are named
CC after the database hits. Genes without significant peptide
CC similarity but with EST similarity are named as 'unknown' proteins.
CC Genes without protein or EST similarity, that are predicted by more
CC than two gene prediction programs over most of their length are
CC annotated as 'hypothetical' proteins. Genes encoding tRNAs are
CC predicted by tRNAscan-SE (Sean Eddy,
CC <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are
CC identified by RepeatMasker (Arian Smit,
CC <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
XX
FH Key Location/Qualifiers
FH source 1. 81020
FH /chromosome="III"
FH /db_xref="taxon:3702"
FH /organism="Arabidopsis thaliana"
FH /cultivar="Columbia"
FH /clone="MGH6"
FH /join(<426..518,645..>1034)
FH /gene="MGH6.1"
FH join(426..518,645..1034)
FH /codon_start=1
FH /db_xref="SPTREMBL:O9M7W6"
FH /note="unknown protein"
FH /protein_id="AAF35940.1"
FH /translation="MMHERPLPANTIPGGPHSNPTTSLHAEVTLCRIFKRTVS
SKRYTDWRELKANGKQOQSNYQAYINFGDNESSTNVNVRREGKNGYERSFQI
QPTPQHQNPILMDTHVDVSFOHFSNDNIHETPTWDELRSVVVEAFPPSFLS"
FH complement(2496..2802)
FH /note="ATTIR16T3A|ATTIR16T3A ATTIR16T3A, a non-autonomous
FH DNA transposon - a consensus."
FH complement(3046..3122)
FH /rpt_family="(GAA)n"
FH join(complement(13663..>13709),complement(13169..13565),
FH complement(<12513..13076))
FH /gene="MGH6.3"
FH join(complement(13663..13709),complement(13169..13565),
FH complement(12513..13076))
FH /codon_start=1
FH /db_xref="SPTREMBL:O9LDD1"
FH /note="hypothetical protein"
FH /gene="MGH6.3"
FH /protein_id="AAF35941.1"
FH /translation="NAVEAHLNLPSSNREMIHPVEASGVVNTOMRYCTVPTNPTV
EQOTSLFPIYNI SPVDRLVHQSMPKTIQSVSSSLTFNSDNVDFLRVPSSKRREES
VYINSAQMOGNKDPDPLFM LQGDLSNVQQHFDIDRLI SNHVERMREIEERKKTQ
GRRIVEAQOGLMKT LRAKDDEINHIGKLNLEEKVKSLCVSNQIWRDVAQSNKATV
ALRNLQYLAVERNREEPTADDAQCCGSDGDEEERMKLAGAQTCKMCR
VGMSCRCGKGEASVLLPCRMCLCSVCGSSLNTPCICKSPKTAASHVNLSS"
FH complement(13950..14024)
FH /rpt_family="(TAAA)n"
FH join(<18851..19090,19255..19381,19476..19614,19697..19785,
FH 19881..>20002)
FH /gene="MGH6.4"
FH join(18851..19090,19255..19381,19476..19614,19697..19785,
FH 19881..20002)
FH /codon_start=1
FH /db_xref="SPTREMBL:O9LDY9"
FH /note="hypothetical protein"
FH /gene="MGH6.4"

FT /protein_id="AAF35942.1"
FT /translation="MASTGLTVAGALLAGDFRLPAVSSSLIPKRTSSSLSCSLNRDLSS
PYNCWRLSRGKILTSLSNKRKFAVGEAEDGFLSNVSEDTDEMFDLFLNKYKRVFKS
TDVKSPTAEYDDDAEALAFALAKVASDKAGDIKVLFPKPLVYTWTFRTTATAFSRP
QIDAIGSRMRDLAEKTKYKVGANGVKPNSMTLLDFGDVVJHLFLPPORTFYNLDFYGN
AMOTELPFEDQSOPRN"
FT join(complement(21254..22367),complement(<20664..21091))
FT /gene="MGH6.5"
FT join(complement(21254..22298),complement(20664..21091))
FT /codon_start=1
FT /db_xref="SPTREMBL:O9M7W5"
FT /note="unknown protein"
FT /gene="MGH6.5"
FT /protein_id="AAF35943.1"
FT /translation="MMAGNOMPTLGRVKVCDLPVSEGLPSDSYKLVATVTLSSLSAQVSA
AIIQFPASDGLALLRSGLSDSARIYFQHDSDYPATNMIHTNDVQWCKTSGYAADPQSQW
EYSEYRPLGTPTSPNSMSEFPAGLDFIALIGKAARVJDAIGFYLNLRSCPTTEILP
NVPLRNCVSSVSVLSCCYARPSFPGAHQHSLEDEOLILYSDHDHLDKSLISFVKSD
KAGLHTRMHGQWILVDYDLGPQEAHVYVPGALYQATAGYVSPAVHRTDLNSLGSISG
RFLAFKLMPSKMTNLSCDSMRAGHGVQFQLPVSDVDFMQHSHSNDLFLNROTLOS
FIVQSDGSMKQLKRRKSDSRCKPLPPLSRKLRLKLEAQRVLKRVQETADKKGILKRF
NLKECENHNVMSPCANIRREIGWPHGVFPVHPHDLPNKAKIGFLETPEPWSETHDM
EFSLSPTAOCNOHVNTNCHSSPNNTPPPFAET"
FT join(complement(24815..>25204),complement(24582..24731),
FT complement(24256..24496),complement(23677..24177),
FT complement(<23184..23578))
FT /gene="MGH6.6"
FT join(complement(24815..25204),complement(24582..24731),
FT complement(24256..24496),complement(23677..24177),
FT complement(23184..23578))
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LDF8"
FT /note="unknown protein"
FT /gene="MGH6.6"
FT /protein_id="AAF35944.1"
FT /translation="MPOGSTGOHCFTAAFSYFWSPTSSRLSNAAEERANYFSNLOKEE
DDDEYSPVSTPEKPGORATLLELMTIRAFHSKMLRCYSLGTAIGFIRRGVLTDP
AIVFVSRKVKHKNWLSPLQCLTALEGAGGIVCDVDVVEFSYFGEPOHQPTPKOTFTTD
IVDLQSDPFIQGSQVSOETCGTIGAIIVRSOTQGVFTVNRHVAVNLDYPSOKM
FHPPLPAGVGLYGEVARATSIITDMLFQIFAGTNPETTFVADGAFIPADYDILSR
VTTSPAGVGEIYGEVARATSIITDMLFQIFAGTNPETTFVADGAFIPADYDILSR
CFUTDLVCGENHRSFPLDGSGLIVMKGEKARPIGITWGCTGSRGLKLVGECIP
ESWTGVDJGLRLTHLQDLITTDEGLKAQVQORAASTTCMSMWADSSPPYVNLKKE
KRSPERLEASLGPLQVQVHIDLEERIEITKGAPSVHQFMPPTFSGQCSASAWPETARE
LVAGFTNGSCDGLCVGLKIGDQKRRRTQVTRNWRPAE"
FT complement(26470..26521)
FT /rpt_family="(GA)n"
FT <26932..>27192
FT /gene="MGH6.7"
FT 26932..27192
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LE44"
FT /note="hypothetical protein"
FT /gene="MGH6.7"
FT /protein_id="AAF35945.1"
FT /translation="MAKNKDDIKYATAQAALSEDAIRVSKYHKTPLLEGCKIAESEPVE
LFSSARIEKKEQSAASGQTOIQRDIDKIKTRTDDSPR"
FT complement(<31715..>32860)
FT /gene="MGH6.8"
FT complement(31715..32860)
FT /codon_start=1
FT /db_xref="SPTREMBL:O9LD39"
FT /note="unknown protein"
FT /gene="MGH6.8"
FT /protein_id="AAF35946.1"
FT /translation="MASGVKNVGTSPSHSWTSKMSLTRESQPLAPALNEPDVDFE
FLEDDPVTMLSADELFSDGLVPLKFTYPEEKPTTSVTVAVKPCRLEMEISGV
DPYLFSPRAPRCTVRWRLELGLARLAKTQOEASASSRSLSSSPNPKTASFRFLNRS
SKTAQOPSHPPKCKSDILESSSTSISSRSLSSSSSSSGHEDLDLPRSLDLDNKPG
TPNPFARSAHHHHHLLRNQPKRPKTRQVDESTESSIESRVTWTPSDPRLNASKI
VFHLERSSSSPCNFTGGPRMKLHHGMRSHSNVRITPVNLVPVSSLSRSGPKSGLEFFG

Query Match

4.5%; Score 74.2; DB 24; Length 81020;

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	1644	100.0	2562	21	AA4249686		Human organic cation
2	441.2	26.8	1906	21	AA4249689		Rat organic cation
3	359.8	21.9	401	21	AA4453112		Human secreted exp
4	295.4	18.0	480	21	AA4242410		Human 5' EST iso
5	66.2	4.0	3580	21	AA4166597		Human secreted pro
6	66.2	4.0	4366	22	AA4575566		Human brain cell s
7	63	3.8	2268	22	AA4263353		P. putida oxygenas
8	49	3.0	1356	22	AA4683133		C glutamitum codin
9	49	3.0	309400	22	AA4685134		C glutamitum codin
10	46.6	2.8	1662	19	AA4495559		Human liver cell c
11	46.6	2.8	1888	19	AA4495558		Human liver cell c

CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.

XX
 SQ Sequence 3580 BP; 734 A; 942 C; 1024 G; 880 T; 0 other;

Query Match 4.0%; Score 66.2; DB 21; Length 3580;
 Best Local Similarity 47.3%; Pred. No. 1.6e-08;
 Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 231 tggtttgaaataatttcagtggaagctgtctgtcttctcaactggcttggtggatggtcga 290
 Db 504 tggccacggcgtctccagtggaacactgtatttggctgtgctggtcgtgctgctga 563
 QY 291 tgcctatggatgatgatcctcagcatcctggcaccacagctgcatgagtgagggct 350
 Db 564 cgggtggaggtcttgggtgggtcttgcgtccagcgcgtgagaagacatgtgct 623
 QY 351 cccaaagctggcagggtggtggtggtggtggtggtggtggtggtggtggtggtggt 410
 Db 624 gtccgactcccaaaaggcatgctggtggtggtggtggtggtggtggtggtggtggtggt 683
 QY 411 cagcgtctgggaaataatctcagaccagtcagcaggaacacagggtcgaagatcagcgt 470
 Db 684 ctctctctgggaggtgctgctgaccggtgggtcggagcagtgctgctcctcgtcgt 743
 QY 471 gctgtggaactctgactatggcctccttagtgcatttgcgcctgtatagcgtggtcct 530
 Db 744 ctccagtaacagcgtctctcctctctctctctctctctctctctctctctctctctct 803
 QY 531 ggtgctccggggtcgtggtggtcggggtcggggtcggggtcggggtcggggtcggggt 587
 Db 804 ctctcgcgcctacttctctgggtgtgggtgtgggtgtgggtgtgggtgtgggtgtgggt 863
 QY 588 gtatgcccagttcctcccatgaaagccagagcgtataatgtatttgcgtgaggtatt 647
 Db 864 ttctccagtttctggccagagaaacagggagcatttggagctggctcgtcgtgtt 923
 QY 648 ctgggcatcgggacaggttccaggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 707
 Db 924 ttggatgattggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 983
 QY 708 ctggcgttggcgtgct 722
 Db 984 gtgagttttcagat 998

RESULT 6
 AAH57556
 ID AAH57556 standard; cDNA: 4366 BP.

XX AC AAH57556;
 XX DT 10-SEP-2001 (first entry)
 XX DE Human brain cell specific cDNA sequence SEQ ID NO:396.
 XX KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle; ss;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas;

us-09-911-667a-3.rng

CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.

XX
 SQ Sequence 3580 BP; 734 A; 942 C; 1024 G; 880 T; 0 other;

Query Match 4.0%; Score 66.2; DB 21; Length 3580;
 Best Local Similarity 47.3%; Pred. No. 1.6e-08;
 Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 231 tggtttgaaataatttcagtggaagctgtctgtcttctcaactggcttggtggatggtcga 290
 Db 504 tggccacggcgtctccagtggaacactgtatttggctgtgctggtcgtgctgctgctga 563
 QY 291 tgcctatggatgatgatcctcagcatcctggcaccacagctgcatgagtgagggct 350
 Db 564 cgggtggaggtcttgggtgggtcttgcgtccagcgcgtgagaagacatgtgct 623
 QY 351 cccaaagctggcagggtggtggtggtggtggtggtggtggtggtggtggtggtggt 410
 Db 624 gtccgactcccaaaaggcatgctggtggtggtggtggtggtggtggtggtggtggtggt 683
 QY 411 cagcgtctgggaaataatctcagaccagtcagcaggaacacagggtcgaagatcagcgt 470
 Db 684 ctctctctgggaggtgctgctgaccggtgggtcggagcagtgctgctcctcgtcgt 743
 QY 471 gctgtggaactctgactatggcctccttagtgcatttgcgcctgtatagcgtggtcct 530
 Db 744 ctccagtaacagcgtctctcctctctctctctctctctctctctctctctctctctct 803
 QY 531 ggtgctccggggtcgtggtggtcggggtcggggtcggggtcggggtcggggtcggggt 587
 Db 804 ctctcgcgcctacttctctgggtgtgggtgtgggtgtgggtgtgggtgtgggtgtgggt 863
 QY 588 gtatgcccagttcctcccatgaaagccagagcgtataatgtatttgcgtgaggtatt 647
 Db 864 ttctccagtttctggccagagaaacagggagcatttggagctggctcgtcgtgtt 923
 QY 648 ctgggcatcgggacaggttccaggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 707
 Db 924 ttggatgattggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 983
 QY 708 ctggcgttggcgtgct 722
 Db 984 gtgagttttcagat 998

RESULT 6
 AAH57556
 ID AAH57556 standard; cDNA: 4366 BP.

XX AC AAH57556;
 XX DT 10-SEP-2001 (first entry)
 XX DE Human brain cell specific cDNA sequence SEQ ID NO:396.
 XX KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle; ss;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas;

us-09-911-667a-3.rng

CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.

XX
 SQ Sequence 3580 BP; 734 A; 942 C; 1024 G; 880 T; 0 other;

Query Match 4.0%; Score 66.2; DB 21; Length 3580;
 Best Local Similarity 47.3%; Pred. No. 1.6e-08;
 Matches 234; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 231 tggtttgaaataatttcagtggaagctgtctgtcttctcaactggcttggtggatggtcga 290
 Db 504 tggccacggcgtctccagtggaacactgtatttggctgtgctggtcgtgctgctgctga 563
 QY 291 tgcctatggatgatgatcctcagcatcctggcaccacagctgcatgagtgagggct 350
 Db 564 cgggtggaggtcttgggtgggtcttgcgtccagcgcgtgagaagacatgtgct 623
 QY 351 cccaaagctggcagggtggtggtggtggtggtggtggtggtggtggtggtggtggt 410
 Db 624 gtccgactcccaaaaggcatgctggtggtggtggtggtggtggtggtggtggtggtggt 683
 QY 411 cagcgtctgggaaataatctcagaccagtcagcaggaacacagggtcgaagatcagcgt 470
 Db 684 ctctctctgggaggtgctgctgaccggtgggtcggagcagtgctgctcctcgtcgt 743
 QY 471 gctgtggaactctgactatggcctccttagtgcatttgcgcctgtatagcgtggtcct 530
 Db 744 ctccagtaacagcgtctctcctctctctctctctctctctctctctctctctctctct 803
 QY 531 ggtgctccggggtcgtggtggtcggggtcggggtcggggtcggggtcggggtcggggt 587
 Db 804 ctctcgcgcctacttctctgggtgtgggtgtgggtgtgggtgtgggtgtgggtgtgggt 863
 QY 588 gtatgcccagttcctcccatgaaagccagagcgtataatgtatttgcgtgaggtatt 647
 Db 864 ttctccagtttctggccagagaaacagggagcatttggagctggctcgtcgtgtt 923
 QY 648 ctgggcatcgggacaggttccaggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 707
 Db 924 ttggatgattggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 983
 QY 708 ctggcgttggcgtgct 722
 Db 984 gtgagttttcagat 998

RESULT 6
 AAH57556
 ID AAH57556 standard; cDNA: 4366 BP.

XX AC AAH57556;
 XX DT 10-SEP-2001 (first entry)
 XX DE Human brain cell specific cDNA sequence SEQ ID NO:396.
 XX KW Human; tissue specific; diagnosis; brain; heart; skeletal muscle; ss;
 KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas;

CC regio-specific oxidations of a wide range of substrates, reactions that
CC are difficult or impossible to do with conventional oxidizing agents.
XX
XX Sequence 2268 BP; 371 A; 713 C; 732 G; 450 T; 2 other;

Query Match 3.8%; Score 63; DB 22; Length 2268;
Best Local Similarity 45.9%; Pred. No. 1.le-07;
Matches 253; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

Qy 233 gcttggaaaattccagtggaagtgtctgttctcactggcttggttgatgagcgtgatg 292
Dy 41 gctttaccctgttccactgagtgatgctgctggtgctgctgctgctgctgctgctg 100
Qy 293 ccatggagatgagatcctcagcactcctcctcagcagcagcagcagcagcagcagc 352
Dy 101 gctatgacctgttcactcagctggtgactgcgctcactcagcagcagcagcagcagc 160
Qy 353 caagctggcaggtggcattgtgactcctcgtggttctttaggcactgatgtccagctcca 412
Dy 161 cccctgtgcaggtggtgcactggcagcactatgcgtgttcggcactgatgttcggcgcc 220
Qy 413 cgtctgtgggaatatctcagaccagtcacggcaggaagaaacaggcgtgaagatcagcgtgc 472
Dy 221 tggcttcggcagctggccagaccgcatcgggcgaagaaggcattgcccattgtttcg 280
Qy 473 tgtgacctgtactatggcactccttagtcatttgcgcccgtgtatagctggatcctgg 532
Dy 281 cctgttctcggggacatctcctaatgctttgcagcaaacagcagcaggttgcca 340
Qy 533 tgcctcggggcctggtggcttcggatcgagagagtt---cccagtcggtgacgctgt 589
Dy 341 tctaccgttccatcgccgctggtgctgtggcgctgagtcaccaacgctgtggcactga 400
Qy 590 atgcagatctctccatgaaagccagagctaaatgtatttctgctgattgaggtattct 649
Dy 401 tgaacgaatacgcacccacagcgcctgcagcagcgtgggtggcagatcatgttcaagtgtt 460
Qy 650 gggccatcggacagtgctcagagtcgctgctggctggtgtgttctgctgacccagcctggct 709
Dy 461 attcgtggcgagcagtgctgcggcaggtgtcggcactcttctcagcgtcgcgctgttgct 520
Qy 710 ggcgtgtgctcactcctcagctgtccgcctcctcctcttcttgcgctgctgttcttct 769
Dy 521 ggaagtcca-gttcttgcgcagcgggtgccactgctgctgttaccggatctctact 580
Qy 770 ggtgcctgaa 780
Dy 581 acctgcctgaa 591

RESULT 8
ID AAH68313
AAH68313 standard; DNA; 1356 BP.

XX AAH68313;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum coding sequence fragment SEQ ID NO: 3348.
DE
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
KW
XX Corynebacterium glutamicum.
OS
XX EP1108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-0127688.
PF
XX 16-DEC-1999; 99JP-0377484.
PR

PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR P-PSDB; AAG93094.
XX

Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 8; SEQ ID NO: 3348; 246pp + Sequence Listing; English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present invention is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 1356 BP; 230 A; 388 C; 383 G; 355 T; 0 other;

Query Match 3.0%; Score 49; DB 22; Length 1356;
Best Local Similarity 45.5%; Pred. No. 0.00083;
Matches 215; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

Qy 289 gatgccatggagatgatgattcctcagcactcctggcaccacagctgcatctgcagtgagg 348
Dy 184 gacggctacgacctcattgtgtacggcagcctccaatcggcctcctggctaaaggatggaa 243
Qy 349 ctcccaagctggcaggtggcattgtgacctggtggtgttctttaggcattgatgtccagc 408
Dy 244 ttaagctctcaacgctgggcaccatcgctccaccgcttcttggcattggcgatcgcc 303
Qy 409 tccacgctctgggaaataatctcagaccagtcagcagcagcagcagcagcagcagcagc 468
Dy 304 gctgtgttcatgtgtcagactgtcagaccgctggcgcgaaagcagcgtgtgattggatcc 363
Qy 469 gtgctgtggaactctgtactatggcactcttagtgcatttgcgcccgtgtatagctggatc 528
Dy 364 gtgctgattctctcttccaccatcgtgtgtcatttgcctccaaacccatgggtgttc 423
Qy 529 ctggtgctccggggcctgggtggcttcggatcgagcagcagcagcagcagcagcagcagc 585
Dy 424 ggcgtttccggtttcatcgcagcgttggtctggcggcgttggtgccttcagttaatgcg 483
Qy 586 ctgtatgccaggttccctccatgaagcagcagcagcagcagcagcagcagcagcagcagc 645
Dy 484 atgacctctgtattgttccacgtaagaccatgtcagcgtggcagcagcagcagcagcagc 543
Qy 646 ttctggccatcgggacaggttctcaggttcgctgctgctgctgctgctgctgctgctgctg 705
Dy 544 ggcgttccatcgggttctctatcgcgctgtgctgcaactgtggtgtgttctctctcctcg 603
Qy 706 ggcgtggcgtgtgctgctcactctcctcagctgtcccgctcctcctctctctctctctct 758
Dy 604 gaggagtgggctggtcgttctcactctcctcagcgttgccttgcctcactggtgtgt 656

RESULT 9

AAH68534	
ID	AAH68534 standard; DNA; 309400 BP.
XX	
AC	AAH68534;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum coding sequence fragment SEQ ID NO: 7069.
XX	
KW	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX	organic acid synthesis; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PF	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484.
PR	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
XX	
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI; 2001-376931/40.
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	-
XX	
PS	Disclosure; SEQ ID NO: 7069; 246pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a nucleic acid described
CC	in the exemplification of the invention.
CC	Note: the sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SQ	Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

Query Match	3.0%	Score 49:	DB 22:	Length 309400;
Best Local Similarity	45.5%	Pred. No. 0.015;		
Matches 215; Conservative	0;	Mismatches 255; Indels	3;	Gaps 1;
QY 289	gatgccatgagatgatctccagcatctggcaccacagctgcattcgagatggagg	348		
Db 227907	gacggctacgacctcatgtgtacggcaccgtccactcggcctggctaaggatggaac	227966		
QY 349	ctcccaagctggcaggctggcattgtgaactcggctggtgtcttttgacgatcatctccagc	408		
Db 227967	ttaagctctgaacgctgggcaccatcggctccaccggctctttggcatggcatcggc	228026		
QY 409	tccacgctctggggaaatactcagaccagctacgcccaggagaaacacaggctgaagatcagc	468		
Db 228027	gctgtgtctcattggttcgacgctgcagaccgctggggccgagaaacagcgggtgattggatcc	228086		
QY 469	gtctgttggagactgtactatgatcattctagtgatcttggcccgctgatagctggatc	528		

Db 228087 gtctgattctctgtcttcaccatcgtctgtgcatttgcgtccaaacccatgggtgttic 228144

Qy 529 ctgdtgtctccggggccctgttggtctcggtgatcggagaggagtccccagtcggtga---cg 585
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228147 ggcgttttcgttttcacgcaggcccttgctctggcgcgcttggtgcttcagttaatgcg 228206
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 586 ctgtatcccaggttctctcccatgaagccagagcataaatgtatttttgcgtattgaggtta 645
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228207 atgacctcttgatttggttccacgtaagaccatgcaagcttgggcaacggtgatgatgtcg 228266
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 646 tctgggccatcgggacagtgttcagagtcgtctcgtctcgttgtctgltgatgccagccclq 705
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228267 gccgttcccacgtggttctatcgcgcgctgtgctgcgactigtgtgttcttctctcg 228336
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 706 ggctggcggttggtctcattctctcacgtgttcccgcctctctctcttlggcgl 758
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228327 gagagtgggcgcttcattcttctcatcgttctctcatcgcttgatccactgtttg 228379

RESULT 10

AAV49559

ID AAV49559 standard; cDNA to mRNA; 1662 BP.

XX AC

XX AAV49559;

XX XX

DT 21-OCT-1998 (first entry)

XX XX

DE Human liver cell clone HP01293 cDNA #2.

XX XX

KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
differentiation; immune system; stimulator; suppressor; regulator;
hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
OS Homo sapiens.
XX XX
FN W09821328-A2.
XX XX
PD 22-MAY-1998.
XX XX
PF 07-NOV-1997; 97WO-JP04056.
XX XX
PR 13-NOV-1996; 96JP-0301429.
XX XX
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
XX XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX XX
DR WPI; 1998-297932/26.
XX XX
PT P-PSDB; AAN64538.
XX XX
Human protein having transmembrane domain - useful for, e.g.
research and nutrition
XX XX
PS Claim 3; Page 120-121; 205pp; English.
XX XX
CC AAV49550-V49559 are cDNA sequences which encode human proteins containing
a transmembrane domain. These proteins can be used for, e.g. research
and nutrition, and may have cytokine and cell
proliferation/differentiation, immune stimulating/suppressing,
hematopoiesis regulating, tissue growth, activin/inhibin,
chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
anti-inflammatory or tumour inhibition activity.
XX XX
SQ Sequence 1662 BP; 312 A; 486 C; 466 G; 398 T; 0 other;

Query Match 2.8%; Score 46.6; DB 19; Length 1662;
Best Local Similarity 47.8%; Pred. No. 0.0045;
Matches 232 Conservative 0; Mismatches 244; Indels 9; Gaps 3

```

Db 1092 caccctgcacatggcgccaccagcggaacacctctacctggatttcttactccgctct 1151
QY 1143 ctctgagttccaggtgctctgtgactctgtgattattgaccgctggtggcgcaagaa 1202
Db 1152 ggtcgaataccggggcccttcagccctcaccctcaccctgacccgctggtggcgcatcta 1211
QY 1203 gaccatggcctgtgctttgtcatcttct---ccttctgcagcctcctgctgtttatctg 1259
Db 1212 cccatggcctgtcaaatlgttggggggcgagcctgctcgtcatgatttttatctc 1271
QY 1260 tgttgaagaatagtctcactcgttactcttca---tttgcagagcgtttatttctg 1316
Db 1272 acctgacctgcactggttaaacatcataatcatgtgtgtggccgaatgggaatcaccat 1331
QY 1317 aggtttcgaagcgcatatgtttacacacctgaggtctaccccaacggaacgcgccct 1376
Db 1332 tgaatacaaatgatctgctggtgaatgctgagctgaccccaaatctgctcaggaacct 1391
QY 1377 cggcctgggcacctgcagcgccatggcaagagtggtgctctcactcctcgtttcatcgc 1436
Db 1392 cggagtgatggtgttctcctcctgtgacataggtgggataatcaccccttcatagt 1451
QY 1437 ccagggtgactggaatcctctgtgtacctgactctggcagtttaccagtggtgctgct 1496
Db 1452 ctccaggtgagggaggtctggcaacgttgcctcctcattttgttaccggtgttgacct 1511
QY 1497 cctggctgacctgacctcctgcttttgcctattgagaccagcgaggaactgcagga 1556
Db 1512 gcttgcgcggagtgacgtactcttctca---gagaccaggggtgcgtttgcaga 1568
QY 1557 gtcca 1561
Db 1569 gacca 1573

```

RESULT 11

AAV49558
ID AAV49558 standard; cDNA to mRNA; 1888 BP.

XX AC AAV49558;

XX DT 21-OCT-1998 (first entry)

XX DE Human liver cell clone HP01293 cDNA #1.

XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 90..1745

XX FT /tag= a

XX FT /product= "transmembrane domain containing protein"

XX PN W09821328-A2.

XX PD 22-MAY-1998.

XX PF 07-NOV-1997; 97WO-JP04056.

XX PR 13-NOV-1996; 96JP-0301429.

XX PA (PROT-) PROTEGENE INC.

XX PA (SAGA) SAGAMI CHEM RES CENTRE.

XX PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;

XX DR WPI; 1998-297932/26.

XX DR P-PSDB; AAW64538.

XX Human protein having transmembrane domain - useful for, e.g.
PT research and nutrition
XX Claim 4; Page 141-143; 205pp; English.
XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
CC a transmembrane domain. These proteins can be used for, e.g. research
CC and nutrition, and may have cytokine and cell
CC proliferation/differentiation, immune stimulating/suppressing,
CC hematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
CC anti-inflammatory or tumour inhibition activity.
XX SQ Sequence 1888 BP; 365 A; 545 C; 520 G; 458 T; 0 other;

Query Match 2.8%; Score 46.6; DB 19; Length 1888;

Best Local Similarity 47.8%; Pred. No. 0.0048;

Matches 232; Conservative 0; Mismatches 244; Indels 9; Gaps 3;

QY 1083 cagcctggcctgcgagtagctgagtaggagattacatgacttgcgtggaccacct 1142

Db 1181 catctgcacatggcgccaccagcggaacctctacctgatttcttactccgctct 1240

QY 1143 ctctgagtttccaggtgctctgtgactctggtgattattaccgcctggcgcaagaa 1202

Db 1241 ggtcgaataccggggcccttcacagccctcaccatgaccgctggcgcatcta 1300

QY 1203 gaccatggcctgctgtttgtcatcttct---ccttctgcagcctcctgctgtttatctg 1259

Db 1301 cccatggcctgtcaaatgttggcgggcgagcctgctcgtcatgttttatctc 1360

QY 1260 tgttgaagaatagtcctcactctgttactcttca---ttgcaagagcgtttattctgg 1316

Db 1361 acctgacvtgcactggttaaacatcataatcatgtgttggcggaatgggaatcaccat 1420

QY 1317 aggtcttccaaagcgcatatgtttacacacctgaggtctaccccaacggaacgacct 1376

Db 1421 tgaatacaaatgatctgctgctggtgaatgctgagctgtaccccaattcgcaggaacct 1480

QY 1377 cggcctgggcacctgcagcgccatggcaagatgggtgctctcactcactcgtttcatcgc 1436

Db 1481 cggagtgatggtgttctcctcctgtgtgacataggtgggataatcaccctcctcatagt 1540

QY 1437 ccagggtgactggaatcctcctgtgtacctgactctggcagtttaccagtgctgctgct 1455

Db 1541 ctccaggtgagggaggtctggaagccttgccctcattttgttgcggtgttgggacct 1601

QY 1497 cctggctgacctggcctcctgcttttggccattgagaccagcgaggaactgcagga 1556

Db 1601 gcttgcgcgggagtgacgtactcttctca---gagaccaggggtgcgtttgcccaga 1657

QY 1557 gtcca 1561

Db 1658 gacca 1662

RESULT 12

AAC93874

ID AAC93874 standard; cDNA; 549 BP.

XX AC AAC93874;

XX DT 19-FEB-2001 (first entry)

XX Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:369.

XX Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;

XX KW flea infestation; vaccine; antiparasitic; therapeutic target;

XX KW diagnosis; detection; ss.

XX OS Ctenocephalides felis.

```
XX PN WO200061621-A2.
XX PD 19-OCT-2000.
XX PF 07-APR-2000; 2000WO-US09437.
XX PR 09-APR-1999; 99US-0128704.
XX PA (HESK-) HESKA CORP.
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX DR WPI; 2000-656323/63.
XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX PT acids useful for the prevention, diagnosis and treatment of flea
XX PT infestations -
XX PS Claim 26; Page 368; 964pp; English.
XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX CC or head and nerve cord (HNC) tissue. The invention also relates to the
XX CC encoded proteins. The invention additionally encompasses expression
XX CC constructs, recombinant viruses and recombinant cells comprising the
XX CC nucleic acids of the invention, recombinant production of the proteins,
XX CC antibodies against the proteins, a method of identifying inhibitors of
XX CC the proteins, and compositions comprising the inhibitors for
XX CC administration to an animal. The nucleic acids, and the proteins they
XX CC encode may be used in the prevention, treatment and diagnosis of diseases
XX CC associated with flea infestations. For example, the nucleic acids may be
XX CC used to produce an HMT or HNC protein according to standard recombinant
XX CC DNA methodology by inserting the nucleic acids into a host cell and
XX CC culturing the cell to express the protein. The HMT and HNC nucleic acids
XX CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX CC and quantitate the presence of cat flea or other homologous nucleic acid
XX CC sequences in samples. They may also be used to study the expression and
XX CC function of the proteins and their role in metabolism. The HMT and HNC
XX CC proteins may be used as antigens in the production of specific
XX CC antibodies, and in assays to identify modulators (agonists and
XX CC antagonists) of HMT and/or HNC protein expression and activity. The
XX CC anti-HMT/HNC protein antibodies and antagonists may also be used to
XX CC downregulate protein expression and activity. The antibodies may also be
XX CC used as diagnostic agents for detecting the presence of flea polypeptides
XX CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX CC present sequence represents a cat flea HMT cDNA of the invention.
XX SQ Sequence 549 BP; 162 A; 83 C; 122 G; 182 T; 0 other;

Query Match 2.8%; Score 46.4; DB 21; Length 549;
Best Local Similarity 48.8%; Pred. No. 0.0029;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy 207 ggtggaagatgcagtggaagccattggttggaaatttcagtggaagctgtctgtctt 266
Db 101 gtttgaagatgcgttggcaactcacaggttttggaaatttcacattcttctgtgctg 160
Qy 267 cactggttgcgttgatggtgcatggaatgagatgatcctcagcatcctggcacc 326
Db 161 gagtgagcgtattagatgttttgatggaactcttggaatgagtttgcgttcc 220
Qy 327 acagctgattgcagtgaggagctcccaagctggcagtggtcattgctgactcgttgg 386
Db 221 ttacgacaattgctctggaattaaacacaaacaaaggaataatttaagcgtatagc 280
Qy 387 cttgttagcagatgctccagctccacgctctctgggaaatactcagaccagtcaggag 446
Db 281 ttctataggtattacaagcagcttccacatttatgggggttttttagccctagagagag 340
Qy 447 gaaacagggtcgaag 462
| | | | |
```

```
Db 341 gcggaagtgattatg 356

RESULT 13
AAC95401
ID AAC95401 standard; cDNA; 1590 BP.
XX AC AAC95401;
XX DT 19-FEB-2001 (first entry)
XX DE Cat flea HMT synaptic vesicle 2B-like cDNA ORF, SEQ ID NO:1904.
XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
XX KW flea infestation; vaccine; antiparasitic; therapeutic target;
XX KW diagnosis; detection; ss.
XX OS Ctenocephalides felis.
XX PN WO200061621-A2.
XX PD 19-OCT-2000.
XX PF 07-APR-2000; 2000WO-US09437.
XX PR 09-APR-1999; 99US-0128704.
XX PA (HESK-) HESKA CORP.
XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX DR WPI; 2000-656323/63.
XX DR P-PSDB; AAB29626.
XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX PT acids useful for the prevention, diagnosis and treatment of flea
XX PT infestations -
XX PS Claim 1; Page 926-917; 964pp; English.
XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX CC or head and nerve cord (HNC) tissue. The invention also relates to the
XX CC encoded proteins. The invention additionally encompasses expression
XX CC constructs, recombinant viruses and recombinant cells comprising the
XX CC nucleic acids of the invention, recombinant production of the proteins,
XX CC antibodies against the proteins, a method of identifying inhibitors of
XX CC the proteins, and compositions comprising the inhibitors for
XX CC administration to an animal. The nucleic acids, and the proteins they
XX CC encode may be used in the prevention, treatment and diagnosis of diseases
XX CC associated with flea infestations. For example, the nucleic acids may be
XX CC used to produce an HMT or HNC protein according to standard recombinant
XX CC DNA methodology by inserting the nucleic acids into a host cell and
XX CC culturing the cell to express the protein. The HMT and HNC nucleic acids
XX CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX CC and quantitate the presence of cat flea or other homologous nucleic acid
XX CC sequences in samples. They may also be used to study the expression and
XX CC function of the proteins and their role in metabolism. The HMT and HNC
XX CC proteins may be used as antigens in the production of specific
XX CC antibodies, and in assays to identify modulators (agonists and
XX CC antagonists) of HMT and/or HNC protein expression and activity. The
XX CC anti-HMT/HNC protein antibodies and antagonists may also be used to
XX CC downregulate protein expression and activity. The antibodies may also be
XX CC used as diagnostic agents for detecting the presence of flea polypeptides
XX CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX CC present sequence represents a cat flea HMT cDNA of the invention.
XX SQ Sequence 1590 BP; 482 A; 252 C; 349 G; 507 T; 0 other;

Query Match 2.8%; Score 46.4; DB 21; Length 1590;
Best Local Similarity 48.8%; Pred. No. 0.0051;
Matches 125; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 02:35:33 ; Search time 137.14 Seconds
(without alignments)
2714.962 Million cell updates/sec

Title: US-09-911-667A-3
Perfect score: 1644
Sequence: 1 atdgagagagactattccca.....cgaactctggctctcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
 - 2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
 - 3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
 - 4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
 - 5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq.*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53.2	3.2	1885	3	US-08-501-572-5
2	53.2	3.2	1885	3	US-09-040-444-5
3	45.2	2.7	7218	1	US-08-232-463-14
4	45	2.7	1896	3	US-08-501-572-6
5	45	2.7	1896	3	US-09-040-444-6
6	38.6	2.3	1882	3	US-08-501-572-4
7	38.6	2.3	1882	3	US-09-040-444-4
8	37.6	2.3	50937	4	US-09-428-517-1
9	36.6	2.2	7218	1	US-08-232-463-14
10	36.4	2.2	1490	4	US-08-964-127-3
11	36.4	2.2	2460	4	US-08-964-127-1
12	35.6	2.2	3489	2	US-08-728-323A-1
13	35.6	2.2	32207	2	US-08-770-379-20
14	35.6	2.2	32207	4	US-08-757-669A-20
15	35.2	2.1	2496	1	US-08-073-384C-2
16	35.2	2.1	2496	1	US-08-234-359A-2
17	35.2	2.1	2496	1	US-08-483-043-2
18	35.2	2.1	2496	1	US-08-481-238-2
19	35.2	2.1	2496	2	US-08-471-066B-2
20	35.2	2.1	2496	2	US-08-484-956-2
21	35.2	2.1	2496	2	US-08-757-653-2
22	35.2	2.1	2496	2	US-08-599-491-2
23	35.2	2.1	2496	2	US-08-756-386-2
24	35.2	2.1	2496	2	US-08-823-516-2
25	35.2	2.1	2496	3	US-08-682-853A-2
26	35.2	2.1	2496	3	US-08-759-038-2
27	35.2	2.1	2496	3	US-08-758-314-2

28	35	2.1	2082	4	US-09-440-325A-2	Sequence 2, Appli
29	34.8	2.1	2102	2	US-08-647-397-1	Sequence 1, Appli
30	34.4	2.1	289	4	US-09-007-005-17	Sequence 17, Appli
31	34.4	2.1	289	4	US-09-244-796-17	Sequence 17, Appli
32	34.2	2.1	2484	4	US-09-276-531-46	Sequence 46, Appli
33	34	2.1	2515	1	US-08-061-465-3	Sequence 3, Appli
34	34	2.1	6828	1	US-08-061-465-1	Sequence 1, Appli
35	34	2.1	6828	1	US-07-928-611-12	Sequence 12, Appli
36	33.6	2.0	803	2	US-08-487-811A-12	Sequence 12, Appli
37	33.6	2.0	803	4	US-09-060-694-12	Sequence 12, Appli
38	33.6	2.0	803	5	PCT-US93-07370-12	Sequence 5, Appli
39	33.6	2.0	1610	1	US-08-056-051-5	Sequence 21, Appli
40	33.6	2.0	1610	1	US-07-928-611-21	Sequence 21, Appli
41	33.6	2.0	1610	2	US-08-487-811A-21	Sequence 21, Appli
42	33.6	2.0	1610	2	US-09-060-694-21	Sequence 21, Appli
43	33.6	2.0	1610	5	PCT-US93-07370-21	Sequence 21, Appli
44	33.4	2.0	4089	1	US-07-908-245-1	Sequence 1, Appli
45	33	2.0	729	4	US-08-998-416-749	Sequence 749, App

ALIGNMENTS

RESULT 1
US-08-501-572-5
; Sequence 5, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Finnegan,Henderson,Farabow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4400
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-501-572-5

Query Match 3.2%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 8.3e-06;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
Qy 1083 cagcctggcctgcgagctacgtgagtgaggaggattcgtgctgtgtggaccacct 1142

Db 1164 CATCTGACATGGCGGCACACAGGGAACCTCTACCTGGATTTCTTTACTTCGCTCT 1223
QY 1143 ctctgagttccaggtgctcttgtagctctgtagattatgacctggtggcgcaagaa 1202
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATACCATGACCGGTGGCGGCATCTA 1283
QY 1203 gaccatggcctgctcttctgcatctctctctcctcctcctcctcctcctcctcct 1262
Db 1284 CCCCATGGCGGTGTAACATTAATCATATGCTGTGGCCGAATGGAATCACCATTGC 1343
QY 1263 tggagaataatgctcactcgttactcttca---ttgcaagagcglttattcttgagg 1319
Db 1344 TGACCTCCACTGGTTAAACATCATATCATGCTGTGGCCGAATGGAATCACCATTGC 1403
QY 1320 ctttcaagcgcatatgtttacacacctgaggtctacacacacacacacacacacac 1379
Db 1404 AATACAAATGATCTGCTGGTGAATGCTGAGCTGTACCCACCATTCGTACGGAACCTCAG 1463
QY 1380 cctgggcaacctgagcgccatggcaagagtggtgctctcctcactcctcctcctcctc 1439
Db 1464 AGTATGCTGTCTCCCTGGCAAGCCTTGCCCTCATTTTGTGGCGGTGTGGCGCTGCT 1523
QY 1440 ggtgatgctggaatcctctgtgtagctctgtagctctgtagctctgtagctctgtagct 1499
Db 1524 CAGGCTGAGCGAGCTGCGCAAGCCTTGCCCTCATTTTGTGGCGGTGTGGCGCTGCT 1583
QY 1500 ggtgacctgacctcctcctgttttggcattgagacacacacacacacacacacacac 1559
Db 1584 TGCCGCGGGAGTGACGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
QY 1560 ca 1561
Db 1641 CA 1642

RESULT 2
US-09-040-444-5
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundenan, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040.444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5
Query Match 3.2%; Score 53.2; DB 3; Length 1885;
Best Local Similarity 47.3%; Pred. No. 8.3e-06;
Matches 228; Conservative 0; Mismatches 248; Indels 6; Gaps 2;
QY 1083 cagctggcctgagtagctac 1142
Db 1164 CATCTGACATGGCGGCACACAGGGAACCTCTACCTGGATTTCTTTACTTCGCTCT 1223
QY 1143 ctctgagttccaggtgctcttgtagctctgtagattatgacctggtggcgcaagaa 1202
Db 1224 GGTGAAATCCCGGGGCTTCATAGCCCTCATACCATGACCGGTGGCGGCATCTA 1283
QY 1203 gaccatggcctgctcttctgcatctctctcctcctcctcctcctcctcctcctcct 1262
Db 1284 CCCCATGGCGGTGTAACATTAATCATATGCTGTGGCCGAATGGAATCACCATTGC 1343
QY 1263 tggagaataatgctcactcgttactcttca---ttgcaagagcglttattcttgagg 1319
Db 1344 TGACCTCCACTGGTTAAACATCATATCATGCTGTGGCCGAATGGAATCACCATTGC 1403
QY 1320 ctttcaagcgcatatgtttacacacctgaggtctacacacacacacacacacacacac 1379
Db 1404 AATACAAATGATCTGCTGGTGAATGCTGAGCTGTACCCACCATTCGTACGGAACCTCAG 1463
QY 1380 cctgggcaacctgagcgccatggcaagagtggtgctctcctcactcctcctcctcctc 1439
Db 1464 AGTATGCTGTCTCCCTGGCAAGCCTTGCCCTCATTTTGTGGCGGTGTGGCGCTGCT 1523
QY 1440 ggtgatgctggaatcctctgtgtagctctgtagctctgtagctctgtagctctgtagct 1499
Db 1524 CAGGCTGAGCGAGCTGCGCAAGCCTTGCCCTCATTTTGTGGCGGTGTGGCGCTGCT 1583
QY 1500 ggtgacctgacctcctcctgttttggcattgagacacacacacacacacacacacac 1559
Db 1584 TGCCGCGGGAGTGACGCTACTTCTTCC---AGAGACCAAGGGGACGCTTTGCCAGAGAC 1640
QY 1560 ca 1561
Db 1641 CA 1642

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463


```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-F15
; US-08-232-463-14

```

Query Match 2.7%; Score 45.2; DB 1; Length 7218;
Best Local Similarity 4.0%; pred. NO. 0.0038;
Matches 14; Conservative 194; Mismatches 142; Indels 0

[illegible]

RESULT 4
US-08-501-572-6

```

: Sequence 6, Application US/08501572
: Patent NO. 6063623
: GENERAL INFORMATION:
: APPLICANT: Koepsell, Hermann
: APPLICANT: Grundeman, Dirk
: APPLICANT: Gorboulev, Valentin
: TITLE OF INVENTION: Transport protein Which Effects The
: TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,
: TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan,Henderson,Parabow,Carrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA

```

```

1  ZIP: 20005-3315
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11  SOFTWARE: Patent In Release #1.0, Version #1.30
12
13  CURRENT APPLICATION DATA:
14
15  APPLICATION NUMBER: US/08/501,572
16
17  FILING DATE:
18
19  CLASSIFICATION: 424
20
21  ATTORNEY/AGENT INFORMATION:
22
23  NAME: Toohey, Kimberlin M
24
25  REGISTRATION NUMBER: 35,391
26
27  REFERENCE/DOCKET NUMBER: 02481.1453-00000
28
29  TELECOMMUNICATION INFORMATION:
30
31  TELEPHONE: (202)408-4000
32
33  TELEFAX: (202)408-4400
34
35  INFORMATION FOR SEQ ID NO: 6:
36
37  SEQUENCE CHARACTERISTICS:
38
39  LENGTH: 1896 base pairs
40
41  TYPE: nucleic acid
42
43  STRANDEDNESS: single
44
45  TOPOLOGY: linear
46
47  MOLECULE TYPE: DNA (genomic)
48
49  J5-08-08-501-572-6

```

Query Match 2.7%; Score 45; DB 3; Length 1896;
Best Local Similarity 47.9%; Pred. No. 0.0021;
Matches 194: Conservative 0; Mismatches 205; Indels

Qy	1106	gtgagaggattacatggaactgctgtgagcaccacctctctgagtttccagggtgctctg	1165
Db	1262	GTGACAATATCTACCTGGATTTCTTCTACTCTGCCTCGTTGAATCCCACTCCCTTCA	1321
Qy	1166	tgactctgtggtattatgacccgctggggcgcaagaagaccatagg-----cctgagctt	1220
Db	1322	TGATCATCTCTCATTTATCGACCGCATCGGACGCGTTACCTTGGCGTGCATCAAAATAGG	1381
Qy	1221	tgtcaatctctcttc,gcagcctcctgctgtttatct-gtgttggagaagaaagtgtctca	1279
Db	1382	TTGCAGGGGAGCGTCTGTGGCCTCAGTTTTATACCTGGTGTATCTACAATGGCTAAAAA	1441
Qy	1280	ctctgttactcttcattgcaagagcgtttattcttggaggctttcaagcgcatatgttt	1339
Db	1442	TTATTATCTCATGCTTTGGGAAGATGGGATCACAAATGGCCTATGAGATAGTCTGCCTGG	1501
Qy	1340	acacactgaagtctacccacgcaacgacggccctcgccctgggcacacclgacagcgca	1399
Db	1502	TCAATGCTGAGCTGTACCCACCATTCATTAGGAATCTTTGGCGTCCACATCTGTTCCTCAA	1561
Qy	1400	tggcaagagtggatgctctcatcactccggttcacgcccagggtgatgctggaatccctctg	1459
Db	1562	TGTTGTGCATTTGTTGGCATTCACGCCAATTCCTGGTCTACCGCTCTACATCAATCTGGC	1621
Qy	1460	tgtacctgactctggcagtttaccagtggcgtgctgctcctggctg	1504
Db	1622	TTGAGCTCCCGCTGATGGTTTCGGCTACTTTGGCTCTGGTTGCTG	1666

DECEMBER 5

RESOLUTION 3
US-09-040-444-6
; Sequence 6, Application US/09040444
: Patent No. 6063766

; PACIFIC NO. 6003700
: GENERAL INFORMATION:

APPLICANT: Koepsell, Hermann

APPLICANT: Grundeman, Dirk

APPLICANT: Gorboulev, Valentin

; TITLE OF INVENTION: Transport I

1. TITLE OF INVENTION: Transport of

; TITLE OF INVENTION: DNA Sequencing

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

Qy	1471	ctggcagattacagtggtgctgctctggctgctgctggcctcctgctttttggccatt	1530
Dy	1526	CT--CATTTTGTGGGGTTTGGGCTGACTCTGGGGCCATGACTCTCTCTCCCA	1582
Qy	1531	gagaccaaagcgcgagactgcagaggtccagccacccggagtgggccagagagatg	1587
Dy	1583	GAGACCAAGGGTGTGCTTTGGCTGAGACTATTGAAGAAGCAGAGAACCTGGGAGG	1639

8 T. J. P. 8

```

RESOLVING
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Belach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakhan
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OL
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of
US-09-428-517-1

```

Query Match	2.3%	Score	37.6;	DB	4;	Length	50937;
Best Local Similarity	54.3%;	Pred. No.	1.8;				
Matches	76;	Conservative	0;	Mismatches	64;	Indels	0;
						Gaps	0;

Qy	71	ggtcagagacacacgctctcaggagagcatgaatccagatttgaaggggtccacgtgg	130
Db <th>37483</th> <th>gggttcaggggttcgcggcgatcgtctcgaaatggcgtgacggtgagcgcgcacatga</th> <th>37542</th>	37483	gggttcaggggttcgcggcgatcgtctcgaaatggcgtgacggtgagcgcgcacatga	37542
Qy <th>131</th> <th>gcctadagcgtctgagctgatgaatgggcagctctgcccagaagatttgcacatccca</th> <th>190</th>	131	gcctadagcgtctgagctgatgaatgggcagctctgcccagaagatttgcacatccca	190
Db <th>37543</th> <th>tcctggacgcgtcagcagctgaggaactcgttcgtctgacgcggaagattggggatcgcag</th> <th>37602</th>	37543	tcctggacgcgtcagcagctgaggaactcgttcgtctgacgcggaagattggggatcgcag	37602
Qy <th>191</th> <th>ccgatgatactttcatggtg</th> <th>210</th>	191	ccgatgatactttcatggtg	210
Db <th>37603</th> <th>ccaaagctttgacacagatg</th> <th>37622</th>	37603	ccaaagctttgacacagatg	37622

9
TJUSP

```

RESUL 9
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
:
: GENERAL INFORMATION:
:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
:
: NUMBER OF SEQUENCES: 52
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
:
: COMPUTER READABLE FORM:

```

```

RESULT      7
US-09-040-444-4
; Sequence 4, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And\or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-4

```

[illegible]

us-09-911-667a-3.rni

Tue Mar 12 15:54:35 2002

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.2%; Score 36.6; DB 1; Length 7218;
Best Local Similarity 4.6%; Pred.No. 1.3;
Matches 15; Conservative 175; Mismatches 139; Indels 0; Gaps 0;

QY 3 ggagaggagctattccagctaaagcagctgcgggtgtgaaattccgtgcacagcgca 62
Db 1370 RR 1311
QY 63 ggtgcaagtcagaggagcagcagcttcagagagatgaagtcagattgaagggt 122
Db 1310 RR 1251
QY 123 ccacgtg jgcctagaggtgtgagctgatgagtgagggcagctgtgccaaaggagttgc 182
Db 1250 RR 1191
QY 183 caatccacagtgatactttcatgtgaaatcagtgaaagcagtcattggcttggaaa 242
Db 1190 RR 1131
QY 243 atttcagtgaagctgtctgtctcactggcttggtgagtgagtgatgcatggagat 302
Db 1130 RR 1071
QY 303 gatgacctcagcatctgcagcaccagc 331
Db 1070 RRRATCGCAAGCTCCTCGACCTGCAGC 1042

RESULT 10
US-08-964-127-3
; Sequence 3, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street

CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,127
FILING DATE: 06-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Crews, Ph.D., L. Lee
REGISTRATION NUMBER: P-43,567
REFERENCE/DOCKET NUMBER: 07334/038001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1490 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 492....1349
US-08-964-127-3

Query Match 2.2%; Score 36.4; DB 4; Length 1490;
Best Local Similarity 47.1%; Pred. No. 0.64;
Matches 112; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 351 cccaagctggcaggtggcattgctgacctcgggtggtctttagcgatgatgtccagctc 410
Db 776 CTTGGCTGGCAGGTGATCTCTGGAGCAGATCTCTTCATCTTGGCTTTGCCCTCCGCTA 835
QY 411 cacgctctgggaaatctctcagaccagtcagcagagaaacagggctgaagatcagcgt 470
Db 836 CTTGTTCTGGGTTACCCCGCAGACAGATTGGCCGCGGATTGTGCTGTGACCTT 895
QY 471 gcttggaactctgtactatgagcatccttagtcatttgcgcgctgtatagctggatcct 530
Db 896 GGGCTGTGGGCCCCCTGTGGAGTAGGAGGGCTGTGCTGAGGCTCTCCACAGGCGTCAT 955
QY 531 ggtgctcggggcgtgggtggtcgtcgagatcgaggaggttccccagtcggtgacgctg 588
Db 956 GGCCTCCGATCTCTTTGGGCTTTCTGCTTGGCGGTGTTGACCTGGGTGCTTACCTG 1013

RESULT 11
US-08-964-127-1
; Sequence 1, Application US/08964127
; Patent No. 6277565
; GENERAL INFORMATION:
; APPLICANT: Grandearl, Andrew David John
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804

STATE: New York
COUNTRY: U.S.A.

```

; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-20

Query Match 2.2%; Score 35.6; DB 2; Length 32207;
Best Local Similarity 51.9%; Pred. No. 5.6;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 76 gagacacacggttcaggagagcatgaagtcacagattgaagggtccacgtgggccta 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20886 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGAG 20827
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 136 gagcgtgtgagctgagatgagtgagggcagctgccccaaaggagtttgcacatccccacgat 195
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20826 GAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20767
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 196 gatactttcatgtgtggaagatgcagtggaagcca 229
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20766 AATGAGGACGAGGAGGATGACGAGGAGGAGGACA 20733

RESULT 15
US-08-073-384C-2
; Sequence 2, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; TITLE OF INVENTION: POLYMERASE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-073-384C-2

```

	Query Match	2.1%	Score 35.2;	DB 1;	Length 2496;
	Best Local Similarity	62.5%;	Pred. No. 1.9;		
	Matches 55;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;
Qy	456 gctaaagatcaagcgtgctgtggactctgtactatggcatccttagtcattlqqcccgct	515 			
Db	1740 GCAGAAATCCCGTGGCACCCCTCTTGGCCACGCGCATCCGCCGAGCCTTCGTGCCGA	1799 			
Qy	516 gtatagtgaalcctgctgcagggc	543 			
Db	1800 GGAGGGTGGGTGCTGGTGGCTTTGGAC	1827 			

Search completed: March 8, 2002, 02:39:31
Job time: 10454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 8, 2002, 00:47:58 ; Search time 3543.72 Seconds
(without alignments)
4985.174 Million cell updates/sec

Title: US-09-911-667A-3
Perfect score: 1644
Sequence: 1 atgagagagactattcca.....cgaactctggtctcaggaa 1644

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	850.6	51.7	972	10	AL523335
2	733.6	44.6	1022	10	AL523336
3	576.6	35.1	778	10	AL529154
C 4	567.6	34.5	642	10	AL563357
5	536.6	32.6	638	11	BG703339
6	496.4	30.2	570	11	BI341271
C 7	463.6	28.2	487	10	AW173250
8	436.8	26.6	578	10	AW663941
9	423.2	25.7	468	11	BF775076
10	399.4	24.3	662	11	BI394002
11	364.8	22.2	371	10	AA349847
12	278.8	17.0	536	10	AW965645

13	270	16.4	377	10	AW345918
14	265.4	16.1	402	10	AA350545
15	255.4	15.5	258	11	F08438
C 16	251.8	15.3	450	11	BI296940
17	246.4	15.0	299	11	T10079
18	244.8	14.9	289	10	AW418091
C 19	229.8	14.0	420	11	BF412623
20	229.2	13.9	355	10	A1372624
C 21	228.4	13.9	366	11	T08271
22	226.6	13.8	418	10	BE097508
C 23	226	13.7	941	10	AU067579
C 24	164.8	10.0	742	10	AU220193
C 25	144.2	8.8	308	11	BE056004
26	142.6	8.7	360	10	BE056418
27	138	8.4	441	11	T08675
C 28	134	8.2	328	11	BI134483
29	124.6	7.6	686	10	BE590388
C 30	122.4	7.4	489	13	AZ979001
C 31	104.6	6.4	996	13	CNS01UGB
C 32	104.6	6.4	1047	13	CNS05LV9
C 33	100.6	6.1	300	11	C57850
34	92.8	5.6	253	10	BB586972
35	92.8	5.6	868	13	CNS04275
36	91.2	5.5	1417	12	AK003981
37	90	5.5	275	10	BB593243
38	87.8	5.3	624	10	AU207190
39	86	5.2	563	11	BF641087
40	84	5.1	584	11	BI367211
41	82.2	5.0	208	10	BB590911
42	80.4	4.9	711	11	BF277036
43	79.8	4.9	377	11	C68924
44	77	4.7	640	10	AU056191
C 45	76.4	4.6	1047	13	CNS05PC2

ALIGNMENTS

RESULT 1
AL523335/c

LOCUS
DEFINITION

AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 3 prime, mRNA sequence.

ACCESSION
VERSION

AL523335
AL523335.1 GI:12786828

KEYWORDS
SOURCE

EST.
human.

ORGANISM
Homo sapiens

REFERENCE
AUTHORS

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization

Unpublished (2001)
JOURNAL
COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1..972
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="CS0DC001YN02"
/clone_lib="LTI_NFL003_NBC3"

/sex="male"
/tissue_type="neuroblastoma cells"

/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life

AL523335 972 bp mRNA EST 13-FEB-2001
AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 3

prime, mRNA sequence.

AL523335.1 GI:12786828

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 972)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization

Unpublished (2001)
Contact: Genoscope

Genoscope - Centre National de Sequencage
Bp 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1..972

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DC001YN02"

/clone_lib="LTI_NFL003_NBC3"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life

Tue Mar 12 15:54:38 2002

```

>Db 590 CAGGTGGCATTCTGACCTCGGTGGTCTTTGTAGGATGATGTCAGCTCCACGCTCTGG 649
Qy 421 ggaatattctcagaccagtagcagcaggaagaaacagggctgaagatcagcgtctgtggact 480
Db 650 GGAATATCTCAGACCAGTACGCGAGGAAACAGGGCTGAAGATCAGCGTCTGTGGACT 709
Qy 481 ctgtactatggratecttagtcatttgccgcgcgtgtatagctgatactcgtgctccgg 540
Db 710 CTGTACTATGGCATCTTATGTCATTTGGCCCGGTGTATAGCTGGATCTGTGCTCCGG 769
Qy 541 ggcctggtggcttcgggacagagagagtcacccagtcggtgacgctgataccagttc 600
Db 770 GGCTGTGGCTTCGGGATCGGAGAGTTCGCCAGTCGTCGACGCTGTATGCCAGTTC 829
Qy 601 ctteccatgaagccagagcctaataatgatttctgctgattgaggtattctggccatcgg 660
Db 830 CTTCCTCATGAWGCCAGAGCTAAATGATTTTGTCTGATTGAGTATTCCTGGCCATTCGGG 889
Qy 661 acagtgtcagagtcctcctggtctggttcgtgattgacccagcctggcgtggtggtgctg 720
Db 890 ACAGTGTTCGAGGTGCTCCTGGCTGTGTCGTGATGCCAGCCTGGGCTGGCGTTGGCTG 949
Qy 721 ctaactctcagctgcccgcctcctcctctcttggcgtgctgctgattctggtgctcgtgaa 780
Db 950 CTCAT-CTCTCAGCTGTC--GCTCCCTCTTTGCCGTCGTGTCTTGGCTGC--TGA 1004
Qy 781 agtcaaggtatgatg 798
Db 1005 AATGCAAGGATGATGTG 1022

RESULT 3
AL529154 AL529154 778 bp mRNA EST 13-FEB-2001
LOCUS AL529154 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL529154
VERSION AL529154.1 GI:12792647
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 778)
AUTHORS Li,W.B., Gruber,C., Jessee,J., Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..778
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 150 a 193 c 254 g 166 t 15 others
ORIGIN

```

```

Query Match 35.1% Score 576.6; DB 10; Length 778;
Best Local Similarity 97.3%; Pred. No. 2e-138;
Matches 581; Conservative 13; Mismatches 2; Indels 1; Gaps 1;
Qy 1 atgagaggaggacttattccagtaagcagctgcccgttggtaaatccctcgcacagg 59
Db 182 ATGAGGAGGAGCTTATTCCAGCTAAGCGCCCGGTTGTGAAATCCCTCCACAGG 241
Qy 60 cgaagtcgaagtcagagagcagcagcgttcaggagagcatgaagtcagatgaagg 119
Db 242 CGAGAGTCGAAGGTCAGAGGAGCAGACGCGCTTCAGGAGAGCATGAAGTCCAGATTGAAG 301
Qy 120 ggtccacatgggcttagaggctgtggagctggatgggagcagctgacccaagagtt 179
Db 302 GGTCCACGCTGGGCTTAGAGGCTGTGGAGCTGGATGGGCGCAGCTGTGCCCAAGCATTT 361
Qy 180 tgcgaatccccccagctgatactttcatggtgaagatcagtggaagcattggctttgg 239
Db 362 TGCCAAATCCCAACCGATGATCTTTTCATGGTGAAGATGACAGTGAAGCATTTGGCTTTGG 421
Qy 240 aaatttcagtggaagctgtgttctcactggtggtcagctgagtgagtgatgcatgga 299
Db 422 AAAATTTTCAGTGAAGCTGTCTTCTCCTCAGCTGGCTTGGCTTGGATGGCTGATGCCATGA 481
Qy 300 gatgatgctcctcagcactcctgacccacagctgcatcagtgagtgagagctcccaagctg 359
Db 482 GATGATGATCTTCAGCATCTCTGGCACACAGCTGCTGTTCCGAGTGGAGGCTCCCAAGCTG 541
Qy 360 gcaggtggcattgctgacctcgtggtctttgtaggcatgattccagctccacagctcgtg 419
Db 542 GCAGGTGGCATTCCTGAMCTCGTGTCTTTTCTAGGCATGANGTCCAGCTCCACGCTCTG 601
Qy 420 gggaaatatctcagacacagtagcagcaggaagaaagggctgaagatcagcgtgctgtggac 479
Db 602 GGGAAAWATYTCAGACAGTACGCGCAGGAGAAACARGGCTTGAAGATCAGGCTGCTGTGGAC 661
Qy 480 tctg...ctatggcattcttagtcatttgcgccgtgtatagctgagctgagtgctcgcg 539
Db 662 TCTKTWCTATGGCATCTTATGTCATTTGCGCCCGTKTWTAGCTGGATCCTGTGTGCTCG 721
Qy 540 gggcctggtgggcttcggagcagcagagagagagagagagagagagagagagagagagag 596
Db 722 GGGCCTGCTGGGCTTCGGGATCGGAGGAGTTCGCCAGTCGGTGGATCCTGTGTGCTCG 778

RESULT 4
AL563357/c 642 bp mRNA EST 16-FEB-2001
LOCUS AL563357 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
DEFINITION prime, mRNA sequence.
ACCESSION AL563357
VERSION AL563357.1 GI:12912671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Li,W.B., Gruber,C., Jessee,J., Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD004YD17"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"

```

BASE COUNT	ORIGIN
149 a	192 c 170 g 117 t 14 others

[illegible]

DB	582	AGTAC	TGAGTGAGGAGAGATWACATGACATGCTGTGTGAGCCACACCTCTGAGGATTCGCG	525
QY	1157	gtgtcctctgtgactctgtggattattgaccgcctctggggcgcaagaagaccattggccctgt	1216	
DB	522	GTGTCTTTGTGACTCTGTGGATTATTGACCGCTTGGGGCGCAAGAAGACCATGTGGCCCTGG	463	
QY	1217	gctttgtcaattctctctctctgcaacctctctgcttttatctgttctgttggaaq-aaatgtg	1275	
DB	462	GCTTTGTGATCTTCTCCCTTCGACGCCMCCNGCTGTTTNNCTGTCTTGGAGAAAATGNG	403	
QY	1276	ctcaactctgttactcttcattcattcgaaagcgctttattctctggaggcttttcgaagcgccatat	1335	

Db	402	CTCACTGTGAAC	TCTTCATGCCAAGACGGTTTATTTCTGGAGGCTTTTCAACGGCGAT	343
Qy	1336	gtttacacactgagtgctacccacgcaacgcggccctcgccctgggaccttcgacg		1395
Db	342	GTTCACACCTGAGGT	TACCCACGCAACGGGGCCCTCGCCTGGGCACCTGCAGC	283
Qy	1396	ggcatggcaagtggggtgtctctcatcctccgttcattcgcccaagtgatgtgaatcc		1455
Db	282	GGCATGGCAAGAGNGGGGCTCTCANTACTCCGTTTCATCGCCAGCGTGATGTGGAATCC		223

[illegible]

DQ		102	GCCAGGATGGTCTCAGCGTTTCACGCTTGCTGTGCATTGGCTTCTG
QY	1636	tctcaggaa	1644
		+ + + + + + + +	
DB	42	TCTCAGGA	34
		RESULT	5
BG703339			
LOCUS	BG703339	638 bp	mRNA
DEFINITION	602685286F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817829 5'		
		EST	07-MAY-2001

mRNA sequence.
 BG703339
 ACCESSION
 BG703339.1
 VERSION
 EST.
 KEYWORDS

SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 638)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsh.r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) -DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) -DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10718 row: 1 column: 22 High quality sequence stop: 638.

FEATURES	source
Location/Qualifiers	1. 638
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_image="4817829"	
/clone_lib="NIH_MGC_95"	
/tissue_type="hippocampus"	
/lab_host="DH10B"	
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaag p) Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."	133 a 149 c 212 g 144 t
BASE COUNT	
ORIGIN	

Query Match	32.6%;	Score 536.6;	DB 11;	Length 638;
Best Local Similarity	98.4%;	Pred. No. 4.le-128;		
Matches 605;	Conservative 0;	Mismatches 4;	Indels 6;	Gaps 6;
Qy	1	atggaggaggaacttattccagctaaggcagctgccggttgtgaattccgtcacacaggc	60	
Db	25	ATGGAGGAGGACTTATTCCAGTAAGGCAGCTGCCGGTTGTGAATAATTCGTGCGCACAGC	84	
Qy	61	gaagagtcaagttcgagagacacacgcgttcacagagagcatgaagtcacagattgaaggg	120	
Db	85	GAGAGTGCAAGCTCAGAGACACACGCGCTTCAGGAGAGCATGAAGTCCAGATTGAAGGG	144	
Qy	121	gtccagctaggacctagaagcgttgagcgtgatgagtgaggcagctgtgcccaagagattt	180	
Db	145	GTCACAGTGGGCTAGAGGCTGTGGAGTGTGATGSGGCGAGCTGTGCCAAGGAGTTTT	204	
Qy	181	gccaatccccaccgatatactttcatggtgaaagatgcagtggaagccattggcttttggga	240	
Db	205	GCCAATCCCACATGATGATACTTTCATGTGGAAGATGCAGTGGAAAGCATTTGGCTTTTGA	264	
Qy	241	a-aatttcagtgaaagctctgttctcactgccttgcttgagctggatggctgatgccatgga	299	
Db	265	ACAATTTCAATGGAAAGCTCTGTCTTCTCACTGGCTTGGCTTGGATGGCTGATGCCATGGA	324	
Qy	300	gatgatgactctcagcatcctctggcaccacagctgcattgcagtggaagctcccacagctg	359	
Db	325	GATGATGATCCTCAGCATCCTGGCACACAGCTGCTATTGCGAGTGGAGGCTCCCCAACGCTG	384	
Qy	360	gcaggtggcattgctgaacctcggcttgcttttaggcattgatgtccagctccacgctctg	419	
Db	385	GCAGGTGGCAATTCTCACCCTCGTGCTCTTTGTAGCGATGATGTCCAGCTCCAGGCTCTG	444	

2000

QY	420	gggaaatactcagaccagctacggca -ggaaaacaggggctgaagatcagcgtcgtcgtgga	478
Db	445	GGGAATAATCTCAGACCACTACGGCAGGGAACAGGCGTGACGATCAGCGTCTGTGGA	504
QY	479	ctctgta-ctatggcatcttagtcatttcgcccgtgatagcttgatcctcgtgctc	537
Db	505	CTCTGTACCTATGGCATCTTATGTCATATGCG -CGGTATAGCTGGATCCTGGTGCTC	563
QY	538	cggggcctggtggcttcgggat -cggagaggttcccc-agtcggtgacgctgatatgcgcg	595
Db	564	CGGGCCTGGTGGCTTCGGGATACGAGGAGTTCGCCAAGTCGGTGACGCTGATGCCG	623
QY	596	agtcctctcccatga	610
Db	624	AGTACCTTCCCATGA	638
RESULT 6			
LOCUS	BI341271	570 bp mRNA	30-JUL-2001
DEFINITION	368657 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.	EST	
ACCESSION	BI341271		
VERSION	BI341271.1	GI:15034560	
KEYWORDS	EST.		
SOURCE	pig.		
ORGANISM	Sus scrofa		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
	1 (bases 1 to 570)		
	Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,		
	Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.		
	and Keete,J.W.		
TITLE	Design and use of two pooled tissue normalized cDNA libraries for		
JOURNAL	EST discovery in swine		
COMMENT	Unpublished (2000)		
	Contact: Smith TPL		
	USDA, ARS, US Meat Animal Research Center		
	PO Box 166, Clay Center, NE 68933-0166, USA		
	Tel: 402 762 4366		
	Fax: 402 762 4390		
	Email: smith@mail.marc.usda.gov		
	Single pass sequencing. Bases called and alt_trimmed with phred		
	v0.980904.e. Vector identified by cross_match with the -minscore 18		
	and -mismatch 12 options.		
	PCR Primers		
	FORWARD: AGGAACAGCTATGACAT		
	BACKWARD: GTTTCCTCCAGTCACGACG		
	Plate: 108 row: D column: 4		
	Seq primer: ATTAGGTGACACTATAG.		
FEATURES	Location/Qualifiers		
source	1..570		
	/organism="Sus scrofa"		
	/db_xref="taxon:9823"		
	/clone_lib="MARC 2P1G"		
	/tissue_type="pooled"		
	/lab_host="DH10B"		
	/note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;		
	Library made from pooled tissue from testis, ovary,		
	endometrium, hypothalamus, pituitary, and placenta."		
BASE COUNT	114 a 147 c 156 g 153 t		
ORIGIN			
	Query Match 30.2%; Score 496.4; DB 11; Length 570;		
	Best Local Similarity 91.9%; Pred. No. 9.8e-118;		
	Matches 524; Conservative 0; Mismatches 46; Indels 0; Gaps 0		
QY	619	gctaaatgattttctgctgattgaggtattctggccatcggaacaggttcagatcgctc	678
Db	1	GCTAAGTGCATTTTCTGCTATTGAGGTGTTCTGGGCCATCGGACCGGTGTCAGGTATC	60
QY	679	ctggctgtgttcgtgatgccagcctggcgtggcttggtgctctcatctcagctgc	738

Db	61	CTGGCTCTGTTTGTGATGCCAGCGCTGGCTGGCGTTGGCTGCTCTCTCTTTTCAGCTGTC	120
Qy	739	cgctctctcttgcgtgctgtgttctgctgctgaaagtgaagtgatgatgtg	798
Db	121	CCGCTCTCTCTTCTGCTGTCTGCTGCTTTTGGCTGCCAGAGTGCAGATATGATGTG	180
Qy	799	ctgtcagggaaccaggaaggaatgccaccttaagaggatagcaactgaacacgga	858
Db	181	CTCTCTGGAACACAGGAAAGGCCATTGCCACCTTAAAGAGGATAGCCAGGAAACCGA	240
Qy	859	gtccccatgcgcgtggggaactcatctccagacaggaagacgagcaaatgag	918
Db	241	GCCCCATGCCCTCTGGGAAGCTCATCTCCAGACAGGAAGACCCAGGCGAAATGAGG	300
Qy	919	gacctttcacaccccatcttagatgagcaacttgcctgtgtgtttatatggtttcc	978
Db	301	GACCTTTTCACACCCCATTTTAGATGGACAACCTTGTGTGCTTATATGCTTTTCC	360
Qy	979	aatgattctcttactacggtttagtttactaccacagaaactcttccagcaggagat	1038
Db	361	AATGCGTTTCTTACTATGGACTAGTTCTGCTCACCACTAGAGCTTCCAGCGCTGGAGAT	420
Qy	1039	gtctgcgcactctccagtcggaagaagctgtagagcgaataatgcgctgctgcgcgag	1098
Db	421	GTCGTGACGATCTCCAGCGGAAGAGCGCTAGAGGCAAAATGCAGCTGCGCTCGCAG	480
Qy	1099	tacctgagtgaggagattacatggaactgtgtgaccacccctctctgagttccaggt	1158
Db	481	TACCTTGAGTAAGGAGGATTACATGACCTGCTGTGGACAGCGCTCTCTGAGTTCCAGGT	540
Qy	1159	gtccttgactctgtgatttatgaccgc	1188
Db	541	GTCTTGTGACTCTCTGGATTATGACCGC	570
RESULT	7		
LOCUS	AW173250/5		
DEFINITION	xj5b11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2663997 3' similar to TR:Q9217 Q9217 SV2 RELATED PROTEIN.		
ACCESSION	AW173250	487 bp	EST
VERSION	AW173250.1	GI:6439198	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 487)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov This clone is available royalty-free through LLNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information. Trace considered overall poor quality Seq primer: -400P from Gibco High quality sequence stop: 1. Location/Qualifiers 1. .487 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2663997" /clone_lib="Soares_NFL_T_GBC_S1" /lab_host="DH10B" /note="Organ: pooled; Vector: pT7T3D-pac (Pharmacia) with a modified polylinker; Site:1: Not 1; Site:2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as		
FEATURES	source		

Bovidae; Bovinae; Bos.
 1 (bases 1 to 468)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chilko-McKown,C.G., Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel.: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCCGAGTCAGCAGC
 Plate: 86 row: E column: 5
 Seq primer: ATTAGGTGCACACTAG.

Location/Organisms
 1. .468
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10a"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
 87 a 126 c 139 g 116 t

BASE COUNT
 ORIGIN

	Query Match	25.7%	Score 423.2;	DB 11;	Length 468;	
	Best Local Similarity	94.0%;	Pred. No. 7.9e-99;	Mismatches 0;	Indels 0;	Gaps 0;
	Matches 440;	Conservative	0;			

```

QY      487  tatggcatcttagtcgatttgcgccgtgtatagctggatccctggtgtccccggggccctg 546
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       1  TAGGGCATCTCAGTGCCTTTCCGCCGTGTACAGCTGGATCTGTGTCTCCGGGCCCTG 60

QY      547  gtggggttcgggatcgaggagggttcccagtcggtgacgtgtatccgaagtctctccc      606
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       61  GTGGGCTTCGGGATCGGAGGGTGCCCCAGTCGGGTACACTGTATGTGTAAGTTCCTTCCC 120

QY      607  atgaagccagagctaaatgtattttgctgtattgaggtattcttgggccatcgggacagtg 666
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       121 ATCAAAAGCCAGAGCTAAATGATATTTTGTGATTGAGGTGTTCTTGGGCATCTGGGACAGTG 180

QY      667  ttcgaggtcgctcctggctgtgttcgtgatgccagcctgggtcgtgcttgtgctcctc     726
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       181 TTCCAGGTCGTCTCGCTGTGTFTTGTGATGCCACCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 240

QY      727  ctctcagctgtcccgcctcctcctctcttccgtgtgtgttcttgcctgcctgaaagtcca 786
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       241 CTCTCAGCTGCCACTCTCTCTCTTTCCGGTCTCTGTGTTTTTGGCTGCCGGAGAGTGCG 300

QY      787  aggtatgatgtgtgttcagggaaccagaaaaggcaatcgccaccttaagaggatatga     846
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       301 AGGTATGATGTGTTCGGGAACACAGGAGAAGGCCATCGCCACCTTAAAGAGAAATAGCC 360

QY      847  actgaataacggagctcccatgcctgtgggaaactcatctctccagacaggaagaccga    906
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       361 ACGGAATAACGGAGCTCCCATGCTCTGGGGAAACTCATCAITTTCCAGACAGGAAGACCGA 420

QY      907  ggcaaaatgagggagccttttcaaccccaatttagatggacaactttg 954
  
```

DB	421	GGCAAAATGAGGACCTTTTCACACCCCATTTAGATGACAACTTG	468
RESULT	10		
LOCUS	BI394002		
DEFINITION	pgpin.pk013.c4 Normalized Chicken Pituitary/Hypothalamus/Pineal Library Gallus gallus cDNA clone pgpin.pk013.c4 5' similar to gb AAC78627.1 (AF060173) SV2 related protein [Rattus norvegicus], mRNA sequence.		
ACCESSION	BI394002		
VERSION	BI394002.1	GI:15087284	
KEYWORDS	EST.		
SOURCE	chicken.		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
AUTHORS	Porter, T.E. and Cogburn, L.A.		
TITLE	ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA library USDA/IAFAS Animal Genome Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Larry A. Cogburn University of Delaware Townsend Hall, Newark, DE 19717, USA Tel: 302-831-1335 Fax: 302-831-2822 Email: cogburn@udel.edu, www.chickest.udel.edu . Location/Qualifiers		
FEATURES	1..662 /organism="Gallus gallus" /strain="Commercial broiler chicken" /db_xref="taxon:9031" /clone_lib="Normalized Chicken Pituitary/Hypothalamus/Pineal Library" /sex="Male and Female" /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland" /dev_stage="Embryonic (d12,d14,d19): post-hatch (w1,w3,w5 ,w7,w9)" /lab_host="E. Coli EMDH10B" /note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end" 160 a 145 c 169 g 177 t 11 others		
BASE COUNT			
ORIGIN			

Query Match	24.3%	Score 399.4	DB 11	Length 662
Best Local Similarity	79.0%	Pred. No. 1.3e-92		
Matches 475	Conservative	0	Mismatches 126	Indels 0
1043	gcgacatctccagctcggaagaaggctgtagagcgcaaatgacgcctgcgcgagatcc	1102		
yy				
1	GCAGCATTTCCAGTAGAAGAAAGAGATTAAAGCAAGTGCAGCCTGACCTGCAGTATC	60		
bb				
1103	tgaatgaagagattacatggacttgcctgtgagcacacctctctgattccaggtgtcc	1162		
yy				
61	TGACAGAGGAAGACTACACTGATCTGCTCTGCACAACCTGTCAAGATTCCAGGTGTGT	120		
bb				
1163	tttgactctgtggattattgaccgcctggggcgcaagaagaccatgcccctgctgttg	1222		
yy				
121	TAGTAACACCTGTGGATTATTGATTCGGATAGGCGCGCAAGAAACCAATGCCCTGTCCCTCT	180		
bb				
1223	tcattcttcctctgcagcctcctgctttatctgtgttggagaagaatgctcactc	1282		
yy				
181	TTGTCTTCTCATTTTCGAGCGCTGCTGCTTTCTCTGTGTTGGAAGAAATGTTCTTACTG	240		
bb				
1283	tgttaactcttcattgcaagaagctttatttcttgaggctttccaagcgcatatggttaca	1342		
yy				
241	TGCTCTCTCTCAATGCAAGAGCTTTTATTTCAGAGAGATTTTTCAGGCTGCTTATGTTTACA	300		
bb				

Query Match	25.7%	Score 423.2	DB 11	Length 468
Best Local Similarity	94.0%	Pred. No. 7	9e-99	
Matches 440	Conservative 0	Mismatches 28	Indels 0	Gaps 0
487	tatggcatccttaagcattgcccgcgtatagctgactctagctccgggcctg	546		
bb				
1	TACGGCATCCCTACGTGGTGTTCGCCCGTGTACAGTGGATCTGCTGCTCCGGGCGCTG	60		
547	gtgggcttcgggacgcgagagagttccccagctcggtgacgctgtatgccgaagtctctccc	606		
Db				
61	GTGGGCTTCGGGATCGGAGGGGGTCCCCAGTCGGTCCACACTGTATGCTGAGTCTCTTCCC	120		
607	atgaagccagagctaaagtattttgctgatgaggtattctgggccatcgggacagtg	666		
121	ATGAAGCCACAGCTAAATGTATTTTCTGATTGAGGTGTTCTTGGCCATCGGGACAGTG	180		
Db				
667	tccgaagtcgctcgtgctgtctgatatccagagctggcgtgctgtgctgctcctc	726		
181	TTGAGAGTGTGCTCGGCTGTGTTTGATGCCACAGCTGGGCTGGGCTGCTGCTCATC	240		
Db				
727	ctctcagctgtccgcgctcctcctcttctggcgtgctgtgttctggctgactgaagtgc	786		
241	CTCTCAGCTGTCCCACTCCTCTCTCTTTTTCGGTCCCTGTGTTTTCGGTGGCGAGATGGC	300		
Db				
787	aggtatgatgtgctgtcagggaaccaggaaaaggcaatcgccaccttaagaggaatagca	846		
301	AGGTATGATGTGCTGTCTCGGCAACCAAGGAAAGGCCATCGCCACCTTAAAGAGAAATAGCC	360		
Db				
847	actgaaaaacggagctcccatcgctcgtgggaacctcatctctccagacagggaagccga	906		
361	ACGGAAAACGGAGCTTCCATGCCCTCTGGGAAACTCATCTATTTCCAGACAGGAACCGA	420		
Db				
907	ggcaaatgagggaaccttttcacacccccattttagatggacaacttgg	954		
QY				

[illegible]

RESULT 12	
AW965645	LOCUS
AW965645	536 bp
DEFINITION	EST
ACCESSION	F0137771P
VERSION	MAGE
KEYWORDS	resesquences, MAGI Homo sapiens cDNA, mRNA sequence.
SOURCE	AW965645.1 GI:8155481
ORGANISM	human.
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 536) Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush

```

FEATURES              Location/Qualifiers
    source              1..536
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="WAGE reserences, WAGI"
                        /note="Vector: pBluescriptSKm"
BASE COUNT            95 a 144 c 174 g 123 t
ORIGIN
Query Match           17.0%      Score 278.8;  DB 10;  Length 536;
Best Local Similarity 99.3%      Pred. No. 1.9e-61;
Matches: 240;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

```

[illegible]

/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue_type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"
BASE COUNT 57 a 55 c 82 g 63 t 1 others
ORIGIN

Query Match 15.5% Score 255.4; DB 11; Length 258;
Best Local Similarity 99.2%; Pred. No. 1.6e-55;
Matches 256; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 80 acgacacgcttcaggagagcatgaagtcacagattgaagggtccacgtggcctagagg 139
Db 1 ACGACACGGCTTCAGGAGAGCATGAAGTCCAGATTGAAGGGGTCCACGTGGCCCTAGAGG 60
QY 140 ctgtgagctggatgagggcagctgtgcccagaggtttgcccattcccaccgatgata 199
Db 61 CTGTGAGCTGGATGATGGGGCAGCTGTGCCAAGAGGTTTGCCCAATCCCACTGATGATA 120
QY 200 ctttcattggtggaagatgcagtggaagccattggtttggaaaaatttcagtgaagcgtg 259
Db 121 CTTTCATGTGTGGAAGATGCAGTGAAGCCATTGGCTTTGGAAAAATTTTCAGTGGAACTGT 180
QY 260 ctgttctcactgcttgcttgatgctgctgacatggagatgatgatccctcagcatcc 319
Db 181 CTGTTCTCACTGGCTTGGCTTGGATGGCTGATGCCATGGNGATGATGATCCTCAGCATCC 240
QY 320 tggcaccacacagctgcatt 337
Db 241 TGGCACCACAGCTGCATT 258

Search completed: March 8, 2002, 00:48:17
Job time: 4640 sec

/db_xref="taxon:9606"
/clone_lib="infant brain"
/sex="female"
/dev_stage="infant"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI"
BASE COUNT 72 a 108 c 129 g 85 t 8 others
ORIGIN

Query Match 16.1%; Score 265.4; DB 10; Length 402;
Best Local Similarity 97.8%; Pred. No. 5e-58;
Matches 269; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1370 gggccctcgtggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1429
Db 23 GGGGCCCTCGGCTGGCCACCTGCGAGCGGCATGGCAAGAGTGGGTGCTCTCATCCTCCGT 82
QY 1430 tcatgccacagtgatgctggaatcctctgtgtacatgactctggtgagctgagcagtgct 1489
Db 83 TCATGCCACAGTGATGCTGGAATCCTCTGTGTACTGTACTGTGGCAGTTTACAGTGGCT 142
QY 1490 gctgctcctgagcctgagcctgctgctgctgctgctgctgctgctgctgctgctgctg 1549
Db 143 GCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 202
QY 1550 tgcagagtcac 1609
Db 203 TGCAGGAGTCCAGCACCGGAGTGGGGCCAGAGATGTGCGCCGAGGAATGCACGGTG 262
QY 1610 cagggttaccaggtcgaactcgtgctcagga 1644
Db 263 CAGGTGTTACCAAGTTCGAACCTCTGCTCTCAGGAA 297

RESULT 15
F08438 258 bp mRNA EST 22-FEB-1995
LOCUS HSC2UB081 normalized infant brain cDNA Homo sapiens cDNA clone
c-zub08, mRNA sequence.
F08438
F08438.1 GI:677005
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 258)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,F.,
Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-zub08
Seq primer: (-21)M13-universal.
Location/Qualifiers
FEATURES
source
1..258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zub08"
/clone_lib="normalized infant brain cDNA"
/sex="Female"

• OM of: US-09-911-667A-2 to: GenEmbl.* out_format : pfs

Date: Mar 8, 2002 3:34 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09911667/runat_07032002_130318_29538/app_query.fasta_1.615
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINWATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09911667.ecgnl.1.6277 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-911-667A-2
Query length: 548
Database: GenEmbl.*
Database sequences: 1472140
Database length: 34134837
Search time (sec): 1970.640000

score_list:

Sequence	Strd	Orig	2Score	Len	Documentation
gb_ro:AF060173	+	2749.00	3428.33	1.1e-182	! AF060173 Rattus norvegicus SV2
gb_in:AC004642	+	984.00	1184.24	1.1e-57	! AC004642 Drosophila melanoga
gb_hg:AC020509	-	984.00	1183.68	1.2e-57	! AC020509 Drosophila melanoga
gb_in:AE003462	-	984.00	1177.90	2.5e-57	! AE003462 Drosophila melanoga
gb_pr:HS0802697	+	734.00	909.82	2.1e-42	! AL359592 Homo sapiens mRNA; ch
gb_pl:AB026645	+	723.50	863.92	7.6e-40	! AB026645 Arabidopsis thaliana
em_pl:AC024128	+	723.50	863.92	7.6e-40	! AC024128 Arabidopsis thaliana
gb_in:AE001863	+	644.50	757.07	6.8e-34	! AE001863 Deinococcus radiodu
gb_in:CEK01F9	-	640.50	772.76	9.1e-35	! Z22175 Caenorhabditis elegans
gb_hg:H0702G05	+	572.00	669.94	4.9e-29	! AL442105 Oryza sativa chromo
gb_ba:AE008141	-	568.50	687.78	4.9e-30	! AE008141 Agrobacterium tumefa
gb_ba:AE007266	-	540.50	652.33	4.7e-28	! AE007266 Shorhizobium melill
gb_ba:AE001804	+	527.50	632.50	5.9e-27	! AE001804 Theriotobium maritima
gb_in:CEZK637	-	511.50	604.99	2.0e-25	! Z11115 Caenorhabditis elegans
gb_pat:AX127153	+	498.50	570.52	1.7e-23	! AX127153 Sequence 7069 from
gb_pat:AX123379	+	493.50	613.10	7.1e-26	! AX123379 Sequence 3295 from Pa
gb_pat:AX066931	+	493.50	612.48	7.7e-26	! AX066931 Sequence 513 from Pat
gb_ba:AB000617	+	491.00	580.71	4.5e-24	! AB000617 Bacillus subtilis ge
gb_ba:BS080002	+	491.00	564.16	3.8e-23	! Z99105 Bacillus subtilis com
gb_ba:AE004767	+	466.50	561.02	5.7e-23	! AE004767 Pseudomonas aerugin
gb_pat:E31259	+	446.50	551.22	2.0e-22	! E31258 Liver-specific organic
gb_pat:E31259	+	446.50	551.22	2.0e-22	! E31259 Liver-specific organic
gb_ro:RATLSTP	+	437.50	539.92	8.9e-22	! D27651 Rattus norvegicus lived
gb_ro:DB3044	+	436.00	536.97	1.2e-21	! DB3044 Rattus norvegicus mRNA
gb_ro:RNOCT2R	+	435.00	535.92	1.4e-21	! Y13154 Rattus norvegicus mRNA
gb_ro:RNOCTK	+	435.00	535.80	1.4e-21	! X98334 R.norvegicus mRNA for d
gb_ro:RNCATTRA	+	431.00	531.92	2.4e-21	! X78855 R.norvegicus mRNA for d
gb_pat:A48157	+	430.00	530.67	2.8e-21	! A48157 Sequence 4 from Patent
gb_om:SSOCT2P	+	429.00	525.31	5.9e-21	! Y09400 S.scrofa mRNA for apica
gb_pr:HS077086	+	427.00	526.98	4.5e-21	! Y07086 Human organic cation tr
gb_pat:A48159	+	426.00	525.60	5.3e-21	! A48159 Sequence 6 from Patent
gb_pat:A48158	+	425.50	525.03	5.7e-21	! A48158 Sequence 5 from Patent
gb_ro:MMU38652	+	425.00	524.41	6.2e-21	! U38652 Mus musculus transmembr
gb_pr:HSOCTL	+	425.00	524.39	6.2e-21	! X98332 H.sapiens mRNA for orga
gb_ro:AF010259	+	423.50	522.60	7.8e-21	! AF010259 Mus musculus organica
gb_ro:MMU6036	+	423.00	520.67	1.0e-20	! A006036 Mus musculus organica
gb_pr:HSOCTK	+	423.00	520.29	1.1e-20	! X98333 H.sapiens mRNA for orga
gb_ro:AF060174	+	420.00	515.19	2.0e-20	! AF060174 Rattus norvegicus syna
gb_ro:RATSV2A	+	413.00	503.43	9.1e-20	! L01788 Rattus norvegicus syna
gb_ro:RATSV2A	+	412.00	501.75	1.1e-19	! L05435 Rattus norvegicus syna
gb_pr:AB018278	+	409.50	496.36	2.3e-19	! AB018278 Homo sapiens mRNA for

gb_om:AF015958 + 409.00 501.07 1.2e-19 2732 ! AF015958 Oryctolagus cunicu
gb_pr:AB018279 + 408.50 496.26 2.3e-19 4353 ! AB018279 Homo sapiens mRNA
gb_pat:AX074148 + 407.00 500.84 1.3e-19 2121 ! AX074148 Sequence 4 from Pa
gb_ro:RATSV2BA + 406.00 494.69 2.8e-19 3660 ! L10362 Rattus norvegicus sy

seq_name: gb_ro:AF060173

seq_documentation_block: 2421 bp mRNA ROD 21-NOV-1998
LOCUS AF060173 Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.
DEFINITION Rattus norvegicus SV2 related protein (SVOP) mRNA, complete cds.
ACCESSION AF060173
VERSION AF060173.1 GI:3901267

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 2421)

AUTHORS Janz, R., Hofmann, K. and Sudhof, T.C.

TITLE SVOP, an evolutionarily conserved synaptic vesicle protein,

SUGGESTS novel transport functions of synaptic vesicles

J. Neurosci. 18 (22), 9269-9281 (1998)

REFERENCE 2 (bases 1 to 2421)

AUTHORS Janz, R., Hofmann, K. and Sudhof, T.C.

TITLE Direct Submission

JOURNAL Submitted (16-APR-1998) Center for Basic Neuroscience, Department

of Molecular Genetics, HHMI, UT Southwestern Medical Center at

Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235, USA

Location/Qualifiers

1. .2421

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/tissue_type="Brain"

gene

/gene="SVOP"

CDS

/note="SVOP"

/note="synaptic vesicle protein; similar to transport

protiens"

/codon_start=1

/product="SV2 related protein"

/protein_id="AAC78627.1"

/db_xref="GI:3901268"

/translation="MEEDLQRLPVVFKRTGESARSEDAAAGEHDVQIEGVVRVG
LEAVELDGAAPKKEFANPTDDTFMVDEDAVEAIGFQWKLSDVLTGLAMADAMEMM
ILSILAPQLHCWRPLPSQVALLTSVFVIGMSSSTLMGNSIDQGRKTLKISVFT
IGTVFEVLAVFVMSLGRWLLLSAAPLLVFCVLPKARAKILLIEFWA
TLKRIATENGAPMPLKLLISROEDRGNMRLTFPHFRTWTLTLMLTFWFSNAPSYGL
VLLTTELFOAGVCSISSKKKAVEAKSLACEYLSKEDYMDLLMTLTLSEFPGLVTLW
VIDRGRKTKMALCFVIFSLKLLICIGRNVLTLILFARAFISGGFQQAAYVTPE
VYPTDRLGLGTCGSMARVGLITPFAQVMSLVTLAVYSGCCLLAALASCFL
PIETGRALQESHREWGEMVGRGTNGVPRSNSSGSE"

BASE COUNT 459 a 658 c 735 g 569 t

ORIGIN

alignment_scores:

Quality: 2749.00 Length: 548
Ratio: 5.044 Gaps: 0
Percent Similarity: 99.453 Percent Identity: 95.985

alignment_block:

US-09-911-667A-2 x AF060173 ..

Align seg 1/1 to: AF060173 from: 1 to: 2421

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
|||||
301 ATGAGGAGGACCTGTTCCAGCTCAGGCAGTTGCCGTGGTGAATCCG 350

```

17 gArgThrGlyCylSerAlaArgSerGluAspThrAlaSerGlyGluH 34
|||||
351 CCGCAGGAGAGAGAGCCAGCGTCAGAGAGAGCGCGCTTCGGGGAA 400
|||||
34 IsGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGlu 50
|||||
401 ATGATGTTTCACATTTAGGGGGTCGGAGTGGGCCCTAGAAAGCTG 450
|||||
51 AspAspGlyAlaIaValProLysGluPheAlaAsnProThrAspThr 67
|||||
451 GATGATGAGAGAGCTGTCCTCCCAAGAGTTTCCCAATCCCACTGAT 500
|||||
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGln 84
|||||
501 TTTTCATGTCGAAGATGCGGTGGAGCCATTTGGGTTCGGAAGATTC 550
|||||
84 rPlsLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||
551 GGAAGCTCTCTGTTCTACCGGCTTGGCTTGGATGGCGAGCCCATGG 600
|||||
101 MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
|||||
601 ATGATGATTCAGCATCTCGCGCTCAGCTGCATCGGAGTGGCGACT 650
|||||
117 uProSorTrpGlnValAlaLeuThrSerValValPheValGlyMetM 134
|||||
651 CCCAGCTGGCAGGTGGCGCTGCTGACCTTCGGTGGTCTTCATGGTAT 700
|||||
134 etSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLys 150
|||||
701 TGTCCAGTCTACGCTCTGGGAAACATCTCGGATCAGTATGGCAGGAA 750
|||||
151 ThrGlyLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeuSe 167
|||||
751 ACAGGCTGAAGATCAGTGTGTTCTGGACCTGTACTACGGCATCTCTAG 800
|||||
167 rAlaPheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeuVal 184
|||||
801 CGCTTTCCGCCAGTGTATAGCTGGATCTCGTGTCTCCGAGGCTCGTGG 850
|||||
184 lPheGlyIleGlyValProGlnSerValThrLeuTrpAlaGluPhe 200
|||||
851 GCTTTGGGATGGAGGGTGGCTGAGTGGTACCTGTACCGCGAGTTC 900
|||||
201 LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
|||||
901 CTCCCCATGAAGCCAGAGCCAAAGTGCATTTGCTGATTGAGGTTTCTG 950
|||||
217 pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetPro 234
|||||
951 GGCCATCGGACCGCTGTTCAGGTGCTTCTGGCTGTGTTGTATGCCCA 1000
|||||
234 eLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
|||||
1001 GCCTGGGCTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1050
|||||
251 PheAlaValLeuCysPheTrpLeuProGluSerAlaArgTrpAspValLe 267
|||||
1051 TTTGCTGTTCTGTTCTGCTGCCAGAGAGTCTAGGTACGATGTGCT 1100
|||||
267 uSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284
|||||
1101 GTCTGGGAACCAAGAAAGGCGATTGCTACCTTAAAGCGGATCGCAACAG 1150
|||||
284 luAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 300
|||||
1151 AAAATGGAGCCCATGCTCTGGGAGGCTCATCTCCAGACAGAA 1200
|||||
301 AspArgGlyLysMetArgPheThrProHisPheArgTrpThr 317
|||||
1201 GACCGAGCAAAATGAGGAGCTTTTCACACCCCACTTTCTGTTGCAAC 1250
|||||
317 rLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTyTrpGlyL 334

```

```

1251 TCTGCTGCTGTTATATGTTCTCCAATGCCCTTCTTATTACGGCT 1300
|||||
334 euValLeuLeuThrThrGluLeuPheGluAlaGlyAspValCysGlyIle 350
|||||
1301 TGGTTCTGCTACACAGAACTTTCAGGCGGAGATGTTTGCAGCATC 1350
|||||
351 SerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTy 367
|||||
1351 TCCAGCCGGAAGAAGCGGTGGAAGCAATGCAGCTGGCTTGTGAGTA 1400
|||||
367 rLeuSerGluGluAspTyMetAspLeuLeuTrpThrLeuSerGluP 384
|||||
1401 CCTCAGCAAGAGGATTACATGACCTGTGTGGACACCCCTGTCTGAGT 1450
|||||
384 heProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLys 400
|||||
1451 TCCAGGCTGTCCTTGTGATCTCTGTGGTTCATCGACCGCTGGCGCAAG 1500
|||||
401 LysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeu 417
|||||
1501 AAGACCATGCTCTGTGTTCTGTCATCTTTTCCCTCTGCAGCCTCTGCT 1550
|||||
417 uPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleAlaAla 434
|||||
1551 GTTCATCTGCATTGGAAAGAAATGTGTAACCTCTTACCTTCTTATTGCAA 1600
|||||
434 rGAlaPheIleSerGlyGlyPheGlnAlaAlaValTrpValThrProGlu 450
|||||
1601 GAGCGTTTATTTCTGAGGCTTCCAAGCAGCCTACGTTTACAGCCTGAG 1650
|||||
451 ValTrpProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMe 467
|||||
1651 GTGTATCCAAGCGGACGAGGCGCTGGGCTGGGACCTGCAGCGCAT 1700
|||||
467 tAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuG 484
|||||
1701 GCGGAGGTGGCGCTCATCTCATCTCCATTCATAGCTCAGGTGATGCTGG 1750
|||||
484 luSerSerValTrpLeuThrLeuAlaValTrpSerGlyCysCysLeuLeu 500
|||||
1751 AATCTTCGTGTACTGACCTGGCGCTACAGTGGCTGCTGCTGCTGCTT 1800
|||||
501 AlAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGlyLe 517
|||||
1801 GCTGCTTGGCTCTCTGCTTCTGCTCCATCGAGACCAAGGCGGAGCACT 1850
|||||
517 uGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgGlyM 534
|||||
1851 GCAGGAGTCCAGCCACCGGAGTGGGCGCAGGATGTTGGCCGAGGA 1900
|||||
534 etHisGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||
1901 CAAACAGCAGCGCTCCCGAGTCCAACTCTGCTCTCAGGAG 1944
|||||

```

seq_name: gb_in:AC004642

```

seq_documentation_block:
LOCUS AC004642 148432 bp DNA INV 01-MAY-1998
DEFINITION Drosophila melanogaster DNA sequence (pIs DS00543 (D193) and
ACCESSION DS02867 (D200)), complete sequence.
AC004642 AC004302 AC003551 AC003552 AC003553 AC003554 AC003555
AC003556 AC003557 AC003558 AC003559 AC003560 AC003561 AC003562
AC003563 AC003564 AC003565 AC003566 AC003567 AC003568
AC003569 AC003570 AC003571 AC003572 AC003573 AC003574
AC003575 AC003576 AC003577 AC003578 AC003579 AC003580 AC003581
AC003582 AC003583 AC003584 AC003585 AC003586 AC003587 AC003588
AC003589 AC003590 AC003591 AC003592
AC004642.1 GI:3097827
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in Sac from p1 clones DS00543
(D193) and DS02867 (D200)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

Muscomorphia: Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 148432)
 Celnikier,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazeej,R.G.,
 Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
 Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
 Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
 Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
 Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
 and Kimmel,B.
 Sequencing of Drosophila chromosome 2R, region 60B1-60B10
 Unpublished (1997)

2 (bases 1 to 148432)
 Celnikier,S.E., Aghavani,A., Arcaina,T.T., Baxter,E., Blazeej,R.G.,
 Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
 Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
 Lomotan,M.A., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M.,
 Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,D., Snir,E.,
 Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L.
 and Kimmel,B.
 Direct Submission
 Submitted (01-MAY-1998) Berkeley Drosophila Genome Project, MS
 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
 Berkeley, CA 94720, US
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site
 (http://fruitfly.berkeley.edu/sequence-archive.html) or send email
 to drosophila@mc.lbl.gov.
 Library locations: 63_6, 83_30.
 Location/Qualifiers
 1..148432
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="60B1-60B10"
 /clone="P15 DS00543 (D193) and DS02867 (U200)"
 /note="D193 extends from bp 1 to bp 83271 and d200 extends
 from bp 81802 to bp 148432."
 BASE COUNT 38713 a 34662 c 35252 g 39805 t

ORIGIN

alignment_scores:
 Quality: 984.00 Length: 605
 Ratio: 2.589 Gaps: 19
 Percent Similarity: 62.810 Percent Identity: 39.835

alignment_block:
 US-09-911-667A-2 x AC004642 ..

Align seg 1/1 to: AC004642 from: 1 to: 148432

41 ValHisValGly.LeuGluAlaValGluLeuAsp...AspGlyAlaVal 56
 ::|
 75088 ATCCACATTGGAGCTGAAGCGGTCTATATAAATCCGATGGTAAAGTGGT 75137

56 alProLysGluPhe.....AlaAsnProThrAspAsp 66
 ::|
 75138 CCGAGAGCAATTCACACATTTCCGGGTAAACAGTGCACCTTTTCACAGAT 75187

67 ThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheG1 83
 ::|
 75188 ACCTACACTGTTACAGGCGCATCAATGCTTCGGTTCGGATGTTTCA 75237

83 nPrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG 100
 ::|
 75238 CQTCAGCTGTCGCTCTGTTGGGGCTCTGCTGGATGTCACCTCCATGG 75287

100 lUmMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArg 116

||||| ::|
 75288 AAATGGCCATCTCAGCATATTGGCCCGCTCTTGTCTCGCAGTGGAT 75337

117 LeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMe 133
 ::|
 75338 GTCACCAAGTTCACAGCATCCGTCACACAGTGGTGTCTTTGGCAT 75387

133 tMetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgL 150
 ||||| ::|
 75388 GATGTTGAGTTCCAGCTTTTGGACACAGCTGAGCAACAGATATGCCGGA 75437

150 ys.....ThrGly.LeuLys..... 154
 ||||| ::|
 75438 ATCCGTAAGACTGACTGTAATCAATAGCTAATAGCTACTAAATGCC 75487

155IleSerValLeuTrpThrLeuTrpThrLeuTrpThrGly 164
 ::|
 75488 ACCATTTTCAGGCCCTCACATTTGTCGGAGTACTACTAGTGTG...TATA 75534

164 lYleLeuSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArg 180
 ::|
 75535 GCATTTCTCAGCTGGTGGCTCCAGTTCAGCTGCTGCTGCTGCTGCTAAGA 75584

181 GlyLeuValGlyPheGlyIleGlyValProGlnSer..... 193
 ||||| ::|
 75585 GGTCTGCTAGGCTTCGCCATCGGCTCGCTTCCACAGTCGTGAGTAATCC 75634

194 194
 75635 ATGGACTAGGAGTACCTCTCCACCTAAACTTCTTTAACCCAAAGAG 75684

194 aThrLeuTrpAlaGluPheLeuProMetLysAlaArgAlaLysCys... 209
 ||||| ::|
 75685 TPACCTTGTAGCTGTTTCAGTGTCTCTGGCCCTTGGAGGCTTCTTTGAA 75734

209 209
 75735 GTATTGATGACGCTGAGTACGTCCACCGTGTCTAAACGTATTATTGA 75784

210IleLeuLeuIleGluVal..PheTrpAlaIleGlyThrValPheGlu 224
 ||||| ::|
 75785 ATATTGTTACTGTTTCAGTGTCTCTGGCCCTTGGAGGCTTCTTTGAA 75834

225 ValValLeuAlaValPheValMetProSerLeuGlyTrpArgTrpLeuLe 241
 ||||| ::|
 75835 GTGGTCTCAGCAGTGGTGTGTATCCCTATTACGGATGGGAGGCTCTCT 75884

241 uIleLeuSerAlaValProLeuLeuPheAlaValLeuCys..... 255
 ::|
 75885 GGCCCTTTCGGCTACACCCCTACTCATTTTACCATCTATCGCCGCTTC 75934

255 255
 75935 GTATTCCTCGCTCTCTTTCAGAAATATCCTTAATCAAAATGATCTTTAAAT 75984

256 ...Phe..TrpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsn 270
 ::|
 75985 CACTACAGTGGCTCAGTGAATCGGCTCGCTATCTTCTTACACCGGTCA 76034

271 GlnGluLysAlaIleAlaThrLeuLysArgIle..... 281
 ::|
 76035 AATGATAAGGCCATCAAGGTTCTGGAACAGGTGAGTGGTAACTTTATGGA 76084

282AlaThrGluAsnGlyA 287
 ::|
 76085 GTGATTATACGCTCTAAAACCTACGATTGTAGATAGCTCACAAACAAG 76134

287 laPro..MetProLeuGlyLysLeuIleIleSerArgGlnGluAspArgL 303
 ||||| ::|
 76135 CGCCACATGCTGCTGGTGGCTGATGGCCGACGACGAGCCCGCTGCGC 76184

303 yLys...MetArgAspLeuPheThrProHisPheArgTrpThrLeuL 319
 ::|

seq_documentation_block:	
LOCUS	AC020509 157851 bp DNA HTG 03-JAN-2000
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AC020509
VERSION	AC020509.1 GI:6664388
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 157851)

194 V 194

194 v 194

194 v 194

Tue Mar 12 15:54:23 2002

107632 ATGGAGTAGGAGGTACCTCTCCACCTAAACCTTTCTTTAAACCCAAAAGAG 107681
194 aThrLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCys... 209
|||||
107682 TTACCTTGTACCTGAGTTCTGCCCAACGACCAAGGAAAGTGGCTC 107731
209 209
107732 GTATTGATGGAGCTGAGTACGTGCCACCGTGTGCTTAAACGTATTATTGA 107781
210lleLeuLeuIleGluValPheTTPAlaIleClyThrValPheGln 224
|||||
107782 ATATATTGACTGGTTTCTAGCTCTCTGGCGCTTCGGAGCTTGCTTTCAA 107831
225 ValValLeuAlaValPheValMetProSerLeuGlyTTPArgTrpLeuLe 241
|||||
107832 TGGTCTAGCACTGGTGTGTATCCCTATTACGGATGGCGAGGTCTCTT 107881
241 uIleLeuSerAlaValProLeuLeuPheAlaValLeuCys..... 255
107882 GGCCCTTTCGGCTACACCCCTACTCATTTTACCATCTCTACGCCGGTTC 107931
255 255
107932 GTATTCTCGCTCTCTTCAGAAATATCCTTAATCAAAATGATCTTTAAAT 107981
256 ...Phe...TTPleuProGluSerAlaArgTyrAspValLeuSerGlyAsn 270
: : : : :
107982 CACTACAGTGGCTCAGTGAATCGGCTCGCTATTTACTTCTTACAACGGTCAC 108031
271 GlnGluLysAlaIleAlaThrLeuLysArgIle..... 281
108032 AATGATAAGCCATCAAGGTCTCGGAACAGGTGAGTGGTAACCTTTATGGA 108081
282AlaThrGluAsnCLyA 287
108082 GTGATTATACCTCTAAACCACTACGATTGTAGATAGCTCACACACACAG 108131
287 laProMetProLeuGlyLysLeuIleIleSerArgGlnGluAspArgGI 303
108132 CGCCACATGTGATGGTGGCTGATGGCGGACGAGCCAGCCAGTTCGCC 108181
303 yLys...MetArgAspLeuPheThrProHisPheArgTrpThrLeuL 319
108182 GGAGAGCTTTATAGATCCCTGCTCAGCCGACGACCTGTACAGACCACCATCC 108231
319 euLeuTTPheIleTPPheSerAsnAlaPheSerTyrTyrGlyLeuVal 335
108232 TGTGTGCTTCTCTGCTGGCTGGCTGGCTTTTGTCTACTACGGCCGTGTG 108281
336 LeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSe 352
108282 TGTGTCACGACCGAGCTGCTGGTGGCC..... 108308
352 tArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuS 369
108309 AGGAACAAGGAGAGCCATCCCAAT.....GAGTGGCTCACCTTCA 108348
369 erGluGluAspTyrMetAspLeuLeuThrThrLeuSerGluPhePro 385
108349 TGACCTCCGACTTATGACCTGTGTGGATCACCTGTCCGAGTTTCT 108398
386 GlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysTh 402
108399 GCAATACTGCTCACCATCAAGGTGGTCAAGCTGTTTGGCAAGAAAGAAC 108448
402 rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheI 419
108449 CATTGTTTCGAGTACCTGGCGCTGGTCTGTGCACACTGTGCTGTGAT... 108496
419 leCysValGlyArgAsnValLeu..... 426
|||||

108497GTCCGTGGAAGGTAGTAGTGGTCACACTCAATTGATCATCTGA 108541
427ThrLeuLe 429
108542 TAATCAATCCAATCCCTCTCCAAATCCCTAGTCGCTTCTCCACTTCCCGTAAC 108591
429 uLeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrV 446
108592 CCTCTTCATTCCACGGCGCATATCCGCATCTTCCAGGCCATTATG 108641
446 alTyrThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGly 462
108642 TGTACACTCCGAAATATACCACTGCTTGTAGATCCGTGGAGTTTCG 108691
463 ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl 479
108692 GGCTGTTCCGTGCTGGCGAGACTCGGCCCATGCTGACCCCTTTCGTGGC 108741
479 aGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerG 496
108742 GCAGGTGCTGATGACTCGTCCAGAAATTCAGCCCATGTCACGTACGAA 108791
496 lYcysCysIleLeuAlaAlaLeuAlaSerCysPheLeuProIleLeuThr 512
108792 TCGTGGCGCTGCTGGCTCCATAGCTCGCTCTTCTGCCAGGGAAC 108841
513 LysGly 514
108842 GTTGGC 108847
seq_name: gb_in:AE003462

seq_documentation_block:

LOCUS Al_11462 30542 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013486038 section 11
of 15, complete sequence.
ACCESSION AE003462 AE002575
VERSION AE003462.1 GI:7291637
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota: Metazoa; Arthropoda: Tracheata; Hexapoda: Insecta;
Pterygota: Neoptera: Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 30542)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dunson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evansell,C., Ferraz,C., Ferrieria,S.,
Fleischmann,W., Glasser,K., Glodek,A., Gong,F., Correll,J.H., Gu,Z.,
Gelbart,W.M., Glasser,K., Godkew,A., Hickey,D., Heiman,T.J.,
Huan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kernison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kuip,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.C.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,K.R., Nix,D., Nixson,K.,
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,

[illegible]


```
237779 TAATCAATCAATCTCTCCTCAATCTAGTCGCTTCTCCACTTCCGTAAC 237730
      :::::
429 uleuPheIleAlaArgAlaPheIleSerGlyPheGlnAlaAlaIleVal 446
      |||||
237729 CCTCTCATTCAGCGCGCAGCATATCCGGCATCTTCCAGGCCATTTATG 237680
      |||||
446 alTyThrProGluValTyThrProThrAlaThrArgAlaLeuGlyLeuGly 462
      |||||
237679 TGTACACTCCGGAATATACCCAGCTGCTTGAGATCCGTCGGAGTTTCG 237630
      |||||
463 ThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAl 479
      |||||
237629 GGCTGTTCCGCTCGCCAGACTCGGCCCATGCTGCGCCCTTTCTGTGGC 237580
      |||||
479 aGlnValMetLeuGluSerSerValTyThrLeuAlaValTyThrSerG 496
      |||||
237579 GCAGGTGCTGATGGACTGCTCAGAAATTCAGGCCATGTCAGGTACGCAA 237530
      |||||
496 lyCysCysLeuLeuAlaAlaLeuAlaSerCysPheIleuProIleGluThr 512
      |||||
237529 TCGTGGGCTGCTGGCTCCATAGCTGCTTCTGCCAGGGAACACT 237480
      |||||
513 LysGly 514
      |||
237479 GTTGGC 237474

seq_name: gb_pr:HSM802697

seq_documentation_block:
LOCUS HSM802697 2104 bp mRNA PRI '8-JUN-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp761H039 (from clone LKfZp761H039).
ACCESSION AL359592
VERSION AL359592.1 GI:8655658
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2104)
Blum,H., Bauersachs,S., Mewes,H.W., Weill,B. and Wiemann,S.
Direct Submission
Submitted (15-JUN-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp761H039) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1..2104
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp761H039"
/clone_lib="761 (synonym: hamy2). Vector pSportII; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="amygdala"
1..437
/gene="DKFZp761H039"
<1..437
/gene="DKFZp761H039"
/note="similarity to SV2 related protein (Rattus
norvegicus)"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAB94878.1"
/db_xref="GI:8655659"

FEATURES
source
gene
CDS

/translation="LCFVIFSFCLLIFCVGRNVLTLLIFARAFISGGFQAAYVYT
PEVYPTATRALGLGTCSMARVAGALITPFAQVMLESSVYLTLAVYSGCCLALALASC
FLPIETKRGRLQESSHREMGQEMVGRMGHAGVTRSNSSGSE"
polyA_signal
polyA_site 2090
BASE COUNT 462 a 557 c 579 g 506 t
ORIGIN
alignment_scores:
Quality: 734.00 Length: 144
Ratio: 5.133 Gaps: 0
Percent Similarity: 99.306 Percent Identity: 99.306
alignment_block:
US-09-911-667A-2 x HSM802697
Align seg 1/1 to: HSM802697 from: 1 to: 2104
405 LeuCysPheValIlePheSerPheCysSerLeuLeuPheIleCysVa 421
|||||
3 CTGTGCTTTGTTCATCTCTCTTCTGCAGGCTCTGCTGTTATCTGTGT 52
|||||
421 lGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleS 438
|||||
53 TGAAGAATAATGTCTCACTCTGTTACTTCTTCATTGCAAGAGCGTTATT 102
|||||
438 erGlyGlyPheGlnAlaAlaIleValTyThrProGluValTyThr 454
|||||
103 CTGGAGGCTTTCAGCGGCATATGTTACACACCTGAGGTCTACCCACG 152
|||||
455 AlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVal 471
|||||
153 GCAACGGGGGCTTCGGCTCGCCACCTGCGCCAGGCGCATGGCAAGAGTGG 202
|||||
471 yAlaLeuIleThrProPheIleAlaGlnValMetLeuGluSerValT 488
|||||
203 TGCTCTCATCACTCCGCTTCATCGCCAGGTGATGCTGGAATCCTCTGTGT 252
|||||
488 yLeuThrLeuAlaValTyThrSerGlyCysCysLeuLeuAlaAlaLeuAla 504
|||||
253 ACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
|||||
505 SerCysPheLeuProIleGluThrLysGlyGlyLeuGlnGluSerSe 521
|||||
303 TCCTGCTTTTCCCATTCAGACCAAGGCGGAGGACTGCAGAGTCCAG 352
|||||
521 rHisArgGluThrGlyGlnGluMetValGlyArgGlyMetHisGlyAlaG 538
|||||
353 CCACCGGGAGTGGGGCCAGGAGATGTCGGTCGAGGAATGCACGCTGCAG 402
|||||
538 lyValThrArgSerAsnSerGlySerGlnGlu 548
|||||
403 GTGTTACCAGGTCCGAACTCTGGCTCTCAGGAA 434

seq_name: gb_pl:AB026645
seq_documentation_block:
LOCUS AB026645 81020 bp DNA PLN 07-MAY-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MGH6,
complete sequence.
ACCESSION AB026645
VERSION AB026645.1 GI:4757401
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui PI
clone:MGH6.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Nakamura,K.
REFERENCE
AUTHORS Nakamura,K.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
```

[illegible]

162 yrTyrGlyIleLeuSerAlaPheAlaProValTySerTrpIleLeuVal 178
|||:::|||||:::||| ||| |||:::||||
67866 TAGCTGGTTCTTGAGCCATTTTCACCTAACTACATGTGTTGCATCATT 67915

179 LeuArgGlyLeuValIGlyPheCylIleGlyValProGlnSerValTh 195
|||:::|||||:::||| ||| |||:::||||
67916 CTTCCGCTGTTTTGGTTGGTCTGGGTGGAGAGCGCTGCTTCCTGCCATTC 67965

195 rLeuTrAlaGluPheLeuProMetLvsAlaArgAlatvScVsiIeLeuL 212

[illegible]

68116	CTAGCTTTTACATCACCATCGCATGCAATCTTTTGCAGTTAGTCAATGCC		68155
233			233
233	oSerLeuGlyTrpArgTrpLeuLeuLeuLeuSerAlaValProLeuLeu.		250
68166	AAGATTGGTTGGAGGTGGTTACTTGCATCTCTCTGGTACCTTCTTCGT		68215

68216 TACTTCTTCCTCATTAGGTGGACATCGAATGCTCCGGTGACTTAAT :
LeucineAlaValLeuLysPheMetIleProThrSerArgAlaGlyAspVal
207

68255 LeuSerGlyAsnGlnGluYLaIleAlaThrLeuLysArgIleAlaArh
:::
:::
:::
:::
:::

68286 CTACAAGGCCGAAGCCGAAGCTCTTCGCATATTAGAAGATTGCGAG
:::
:::
:::
:::
:::

293 TGIUASHNRYALPROMELGCTOLEUGLY..... 292
 ||| :::: |||| |
 68316 GATGAACAAAACCACACTCCTCGGGTGTTCAGATTCCAGAGTTGGAGA 68365
 292 292
 68366 CCGAGCTGGAAGAGAAATAAAAAACATTTCCAACAGAAAAAATCATCTTCTC 68415

	LysLeullellese	297
295	: :	
68416	AAGCTGGAGAAGTGGGACGCCGTTCCTGTCTAAGATTGTAATAA	68465
297	rarglncgluaasparggly...LysmetArgAspLuephetrProHisP	313
	: : : : :	
68466	AGCGTAAAGGAACCAGGTTCTTCATTACTAGCTCTTCTATCCCAAGC	68515

68516 TAATGAACGGACTTGTCTACTATGGTTGTCCTCTTGGAAATCGCTTT 68565
:: :::::::::::::::::::: ::::::::::::::::::::
330 SerTyrtyrGlyLeuValLeuLeuthrThrGluLeuPheGlnAlaGlyAsn 346
:: :::::::::::::::::::: ::::::::::::::::::::

68566 GCTTACTATGGCGTTGTCCTTCTCACCAAGCTGAACAATTCCCATAA 68615

[illegible]

```
361 .....CysSerLeu..... 363
68662 AACTACAGACGCTTTTCATTGCCAGTTTCGCAGGTACATCAATATATT 68711
364 .....AlaCysGluTyrLeuSerGluGluAs 372
68712 CTGTTATCAAGAAATCAAGACGTTGTACTTCTTCAACAATAAGAAA 68761
372 pTyrMetAspLeuLeuTrp.....ThrThrLeuSerGluPheProGly 387
68762 AAGTCATAATCTTATTTTGTATCTGTGGCGCTCAGAGTTTCCAGGTC 68811
387 aLeuValThrLeuTrrPillelleAspArgLeuGlyArgLysLysThrMet 403
68812 TACTGATATCAGCAGCAATGGTGATAGACTTGGACGAAAGCTCAATG 68861
404 AlaLeuCysPheValIlePheSerPheCysSerLeuLeuPheIleLeu 420
68862 GCA.....TCAATGTGTTCACATGTTCATTTTCTCTCTCCCTCTGTT 68905
420 sValGlyArgAsn.....ValLeuThrLeuLeuPheIleAlaArgA 435
68906 ATCCATCAGTCTCCATTCATACACCGCTTCTCTTCGGCGCGCGTA 68955
435 laPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluVal 451
68956 TCTGCATCTCGGCGAGCTTTTCACAGTGTGTTCATATATATGCTC 69005
452 TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAl 468
69006 TATCCACGCGAGTGCAGACACCGAGTAGGAGTAGGAGCTCGTAGG 69055
468 aArgValGlyAlaLeuThrProPheIleAlaGlnValMetLeuGluS 485
69056 AAGAAATTGGAGGAATCTTATGTCCACTCGTAGCC...GTGGATTAGTTC 69102
485 erSerValTyrLeuThrLeuAlaVal.....TyrSerGlyCysCysLeu 499
69103 ACGATGCCATCAGACGATCGCTGCTCTCTGTTGAGGTGTGATCTT 69152
500 LeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGly 516
69153 GTGTGAGGAATTTGCGTTTGTCTCTTCCCTTTCGACAGCAAGTGGTCTGTA 69202
516 yLeuGlnGluSer 520
69203 CCTGACAGATAGT 69215
seq_name: em_pl:AC024128
seq_documentation_block:
ID AC024128 standard; DNA; PLN; 81020 BP.
XX AC024128;
AC AC024128;
XX AC024128.1
XX 02-MAR-2000 (Rel. 62, Created)
XX 02-MAR-2000 (Rel. 62, Last updated, Version 1)
DE Arabidopsis thaliana chromosome III p1 MGH6 genomic sequence, complete
DE sequence.
XX HTG.
XX Arabidopsis thaliana (thale cress)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
XX eurosids II; Brassicales; Brassicaceae; Arabidopsi.
XX [1]
RN 1-81020
RP
```

```
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA Matti R., Rønning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III p1 MGH6 genomic sequence";
XX Unpublished.
XX [2]
RN 1-81020
RA Lin X.;
RT Submitted (24-FEB-2000) to the EMBL/GenBank/DBJ databases.
RL The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD
RL 20850, USA
XX SPTREMBL; Q9LD39; Q9LD39.
DR SPTREMBL; Q9LDB0; Q9LDB0.
DR SPTREMBL; Q9LDD1; Q9LDD1.
DR SPTREMBL; Q9LDF8; Q9LDF8.
DR SPTREMBL; Q9LDK4; Q9LDK4.
DR SPTREMBL; Q9LDM2; Q9LDM2.
DR SPTREMBL; Q9LDS0; Q9LDS0.
DR SPTREMBL; Q9LDY9; Q9LDY9.
DR SPTREMBL; Q9LDZ8; Q9LDZ8.
DR SPTREMBL; Q9LE42; Q9LE42.
DR SPTREMBL; Q9LE44; Q9LE44.
DR SPTREMBL; Q9M7W1; Q9M7W1.
DR SPTREMBL; Q9M7W2; Q9M7W2.
DR SPTREMBL; Q9M7W3; Q9M7W3.
DR SPTREMBL; Q9M7W4; Q9M7W4.
DR SPTREMBL; Q9M7W5; Q9M7W5.
DR SPTREMBL; Q9M7W6; Q9M7W6.
XX Address all correspondence to:atetigr.org
CC P1 clone MGH6 is from Arabidopsis chromosome III and is near the
CC molecular marker.
CC The orientation of the sequence is from SP6 to T7 end of the P1
CC clone.
CC Genes were identified by a combination of three methods: Gene
CC prediction programs including GRATL (available by anonymous ftp
CC from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
CC Washington), Genscan (Chris Burge,
CC http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
CC (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
CC complete sequence against a peptide database and the Arabidopsis
CC EST database at TIGR (http://www.tigr.org/tdb/at.html).
CC Annotated genes are named to indicate the level of evidence for
CC their annotation. Genes with similarity to other proteins are named
CC after the database hits. Genes without significant peptide
CC similarity but with EST similarity are named as 'unknown' proteins.
CC Genes without protein or EST similarity, that are predicted by more
CC than two gene prediction programs over most of their length are
CC annotated as 'hypothetical' proteins. Genes encoding tRNAs are
CC predicted by tRNAscan-SE (Sean Eddy,
CC http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
CC identified by RepeatMasker (Arian Smit,
CC http://ftp.genome.washington.edu/RM/RepeatMasker.html).
XX Key Location/Qualifiers
FH 1. 81020
FT source
FT /chromosome="III"
FT /db_xref="taxon:3702"
FT /organism="Arabidopsis thaliana"
FT /cultivar="Columbia"
FT /clone="MGH6"
FT mRNA
FT join(<426..518,645..>1034)
FT /gene="MGH6.1"
FT CDS
FT join(426..518,645..1034)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9M7W6"
FT /note="unknown protein"
FT /gene="MGH6.1"
FT /protein_id="AAF35940.1"
FT FT
```

```
FT FT /translation="MMHEFRILPTANDTIPGGPTHNSNPPTTSLHAEVWTLRIKFKRTVS
FT FT SKKYPPDRRELANGRKQOOSNYEAVINFGDNESSTNNMNVNRECKGNVRSVFUL
FT FT QOTPYQHONQPLMTTTHVDQSHFNSNDNIHETVETWPDELRSVVEFAFPSPFLS"
FT FT complement(2496..2802)
FT FT /note="ATTIR1673A1ATTIR1673A ATTIR1673A , a non-autonomous
FT FT DNA transposon - a consensus.."
FT FT complement(3046..3122)
FT FT /rpt_family="(CAA)n"
FT FT join(complement(13663..13709),complement(13169..13565),
FT FT complement(12513..13076))
FT FT /gene="MGH6.3"
FT FT join(complement(13663..13709),complement(13169..13565),
FT FT complement(12513..13076))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDD1"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.3"
FT FT /protein_id="AAF35941.1"
FT FT /translation="MAVEAHHLNPLFSSNREMIHPVEASGVVYNTOMRYGTVPTNPV
FT FT ECOTSLFNPINISPDRLVHQSMKPTIQSVDSLLTFNSDNNVDFLRPVSSRRKREES
FT FT VVILNSAYWQIQKNPTDPLMFLGODLSNVQOHFEDIDRLISNHVHRMMEIEEKKKTO
FT FT GRRIEAVQGLMKTLRAKDEINHLGKLNLEEKVKSICVENOIWRDVAQSNKATVN
FT FT ALKNSIQVLAVERNRWEEPTTVADDAQSCGSDGSEERKWLAGEAODTKMCR
FT FT VGMCMRSCGKEASVLLPCRHMCILCSVCGSSLNFCPCICKSPKTASLHVNLSS"
FT FT complement(13950..14024)
FT FT /rpt_family="(TAAA)n"
FT FT join(<18851..19090,19255..19381,19476..19614,19697..19785,
FT FT 19881..>20002)
FT FT /gene="MGH6.4"
FT FT join(18851..19090,19255..19381,19476..19614,19697..19785,
FT FT 19881..20002)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDY9"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.4"
FT FT /protein_id="AAF35942.1"
FT FT /translation="MASSGLTVAGALLAGDFRLPAVSSLIIPKHTSSSLCSLNRDLSS
FT FT PYNKWRLSRGKILTLSSNRKFAVEKAEDEGLNSVSDTEDMFDLLEKYGKVFRS
FT FT TYKVSPTAEVDDDAELAFELAKVASDVYKAGDIKLVFKVPLVYVTRFFIATAFSRP
FT FT QIDAIGRMRLDLAEKRYGVANGDVKPNSTWLLDFGVVTHLFLPQRTFYNLDFYGN
FT FT AMOIELPFEDOSOPRN"
FT FT join(complement(21254..22367),complement(<20664..21091))
FT FT /gene="MGH6.5"
FT FT join(complement(21254..22298),complement(20664..21091))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9M7M5"
FT FT /note="unknown protein"
FT FT /gene="MGH6.5"
FT FT /protein_id="AAF35943.1"
FT FT /translation="MMAGNCMPITLGRVKVCDLVPSECLPSDSYKLAVTTLQSIAQYSA
FT FT AIIQFPASDGLRSGLDSRLYFHQRDSYPATNNMHTNDSEWCKTSGYVADPQSQW
FT FT ESEYRPLPTPSPNSMEFPFAGLPDIFALLGKAARVLDLAI GFYLNLRSCPFTEILD
FT FT NVPLRNCVSSVSLVCCYARPFHGAQHSLTEDEOLILYSDHDHOLDKSLISPVKSD
FT FT KAGLHIRMHGMWILVDVLDGPQEAIVYGLAQYTAGVYSPAVHRTDLSLSPFVKS
FT FT RFLAFKLPMTNLSLCSSEMRAAGHVEAQFOLPVQSDVDFMQRSHNSDLFNRQTLQS
FT FT FTVPOSQDGMKOLKRRKSDSRCKPLPPSKRLLEAORVLEKERVIEADKGIKILRFC
FT FT NLECSLETAGNQHVINCNSHSPNPTPPFAET"
FT FT join(complement(24815..25204),complement(24731,
FT FT complement(24256..24496),complement(23677..24177),
FT FT complement(<23184..23578))
FT FT /gene="MGH6.6"
FT FT join(complement(24815..25204),complement(24582..24731),
FT FT complement(24256..24496),complement(23677..24177),
FT FT complement(23184..23578))
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LDF8"
FT FT /note="unknown protein"
FT FT /gene="MGH6.6"
FT FT /protein_id="AAF35944.1"
FT FT /translation="MPOYSGTGOHCETAAYSFSWPTSSRLSNAEERANVFSNLOKEE
FT FT DDDVSEFPEVPSTPKQORATTLLELMTIRAFHSMKLRCLYSLGTAGFRIRRGVLTIDP
FT FT AIIVFSRKRKHQWLSPLQCLPTALEGAGGIMCDVVEFSYFEGPDHPQTPKQTFITD
```

```
FT FT IVDHLQSDPFI GSGSVASQETCCTLCATIVRSOTGRQGVFTNHRHAVNLDPYSOKM
FT FT FHPPLPAGPVI LCAERATSFITDDDLGATFAGTNPETFVRADCAFI PRADDDYDLNR
FT FT VITSYGVGEIVGKATELOSPGSLVKQVAVRSGSLTGTGLV LAYALAYNDERGV
FT FT CSLTDFLVVGENHRSFPFDEGDSGLIVMKGEKARP IGI IGGTGSGRKLKLVGCECP
FT FT EFWTGVDLGLTHLQDLITTDGLKAAVOEAASTTGMSSHVAUSSPPYVNIKKE
FT FT KRSPKEKLEASGLPQVOHIDLEERIEIKGGAPSVEHOFMPTFGCGCSASAMPETARED
FT FT LVAGFTNCSGQDLCVGLRGDDGAKRRRTQVYNERMPAE"
FT FT complement(26470..26521)
FT FT /rpt_family="(GA)n"
FT FT <26932..>27192
FT FT /gene="MGH6.7"
FT FT 26932..27192
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LE44"
FT FT /note="hypothetical protein"
FT FT /gene="MGH6.7"
FT FT /protein_id="AAF35945.1"
FT FT /translation="MAKNKDDIKYATAOAKLSEDEAIRVYKRGTPLEGKIAESEPVE
FT FT LFSSAQRLEKKEQSAASGDOTQIQDIKDKGTRTDDSPR"
FT FT complement(<31715..>32860)
FT FT /gene="MGH6.8"
FT FT complement(31715..32860)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:O9LD39"
FT FT /note="unknown protein"
FT FT /gene="MGH6.8"
FT FT /protein_id="AAF35946.1"
FT FT /translation="MASGCVKNVGTSPSHSTSSKMSLTRESQPLAPALENEDPVDDFE
FT FT FLEDPVTMLSADELFSDGLKLFSGVTYPEEKPTITSVHTAVAPCRRLMEISGVV
FT FT DYLFSPRAPRCTVRWRELLGLRLAKTOQEASASSSSRLSSSSPNPKTASFRHLNRS
FT FT SKSTAQOPSHPPGKSDILESSSTSISSSRLSSSSSSGHELDLDRSLDLDNKPG
FT FT TPNPFAFSFAHHHHLRNONPKPRRHTQVDESTESIESKVMYVTAOSPRLNASGKI
FT FT VFHGLERSSSPNTGGPRMKLIHGMPSHSANVRITPVLNVVPSLSRSGPKSGLFFG
```

alignment_scores:

```
Quality: 723.50 Length: 672
Ratio: 2.055 Gaps: 16
Percent Similarity: 52.361 Percent Identity: 28.423

alignment_block:
US-09-911-667A-2 x AC024128 ..

Align seg 1/1 to: AC024128 from: 1 to: 81020

49 GluLeuAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAs 65
|||||
67249 GAAATGGCGATGGG.....AA 67265

65 pAspThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysP 82
||| |||:|||||: |||:|||||:|||||
67266 CACAAGATTCACGGTGGATGAAGCTCTTGTGCGATGGGATTTGGCAAGT 67315

82 heGlnTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAla 98
||||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||
67316 TCCAAATTTACGCTCTCTTACGCTGGAATGGGTGGGTTCACAGGCC 67365

99 MetGluMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTr 115
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
67366 ATGGAGATGATGCTCTCTCTTCTGCTGGACCTGCTGCTCAATCTCTCTG 67415

115 pArgLeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValG 132
||||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||
67416 GAATCTCTCTGACGCCAGGAAGCTTAATAACTAGCGTGGTTCGTCG 67465

132 LysMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGly 148
||||| :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||
67466 GCATGCTCATTTGGTCTACTCTTGGGGCATGTTTCTGATAACATGCC 67515

149 ArgLys.Thr..... 151
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||
67516 CGCAGTAGCTTTTATCATACTTGTGTTGTTCTTGTATATAAGTTTCTTGTA 67565

151 ..... 151
```

```

67566 CCTCAAGTAGAAGGGGACGGCTGAATTGTTCCGTAATAATGTTTCATATCT 67615
151 ..... 151
67616 TATGAGCTCTTGAATTGAATGATCTCGACCTTGTCTAAAGCTGAATTC 67665
151 ..... 151
67666 AGACTGCTTTAGCTTATGACCATCTTGTGTTTCATCCGATTTCTGTGTCT 67715
151 ..... 151
67716 TGTGTCAATGGTTCATCAATCAATGCTAAATGGACTTGTGTTGGGTCAT 67765
151 ..... 151
67766 ATTCCTCTGTTTATGATGATGCTGTTGCTTCAATTTGATTTGATGTTGTG 67815
152 ..... GlyLeuLysIleSerValLeuTrpThrLeuT 162
67816 CCTTGAATTCCTCAGAAAGGGTTTATCACTACCTCGGTGGTGACTTTTG 67865
162 yTrpMetAspLeuLeuTrp.....ThrThrLeuSerGluPheProGlyVal 178
67866 TAGCTCTTCTTGAGCGCATTTTACCTACATGATGTTGTTGATCAT 67915
179 LeuArgGlyLeuValGlyPheGlyValProGlnSerValTh 195
67916 CTTTCGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 67965
195 rLeuValAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuL 212
67966 TTGGTATCTGATTCATACCTGCACCAAGCGGAGGACTTGGATGCTGTG 68015
212 euLeuGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla 228
68016 TTTTCTCAGCTTTCTGAGCGGTGGAGCAATATTGAGGCTTCCCTGTCT 68065
229 .....Val...Phe..... 230
68066 TGGGTATACTCATGCTTCTTCTGTTTATTTTCCAGTATAATTTGGACG 68115
231 .....ValMetPr 233
68116 CTAGCTTTTTCATCATCACCATGCAATGCAATCTTTTTCAGTGTAGTCATGCC 68165
233 oSerLeuGlyTrpArgTrpLeuLeuLeuSerAlaValProLeuLeuL 250
68166 AAGATTGGGTGGAGTGGTGTACTTGCATTCCTTCGGTACCTTCTCGT 68215
250 euPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTrpAspVal 266
68216 TACTTCTCTGTTCTATAGTGGACATCTCAGTCTCCTCGGTACTTAAT 68265
267 LeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaTh 283
68266 CTACAAGCCGGAAGCGAAGCTCTTGCCATATTAGAGAAGATTGCGAG 68315
283 rGluAsnGlyAlaProMetProLeuGly..... 292
68316 GATGAACAAACCACTACCTCGGCTGTTCTCAGTTCAGAGTTGGAGA 68365
292 ..... 292
68366 CCGACTGGAAGACAATAAACAATTCACAGAAATACTCATCTTCTC 68415
293 .....LysLeuIleIleSe 297
68416 AAGCGTGGAGAAGTGGAGAAGCGTTGCTGTCTAAGATTGTACTAAA 68465
297 rArgGlnGluAspArgGly...LysMetArgAspLeuPheThrProHisP 313
; ..... 111 ..... 111 ..... 111 ..... 111 ;

```

seq_name: gb_ba:AE001863

seq_documentation_block: 198677 bp DNA BCT 22-NOV-1999
LOCUS AE001863
DEFINITION Deinococcus radiodurans R1 section 2 of 2 of the complete
chromosome 2.
ACCESSION AE001863
VERSION AE001863.1 GI:6460670
KEYWORDS Deinococcus radiodurans.
SOURCE Deinococcus radiodurans.
ORGANISM Deinococcus radiodurans.

REFERENCE AUTHORS	Bacteria: Thermus/Deinococcus group; Deinococcales; Deinococcus. 1 (bases 1 to 198677) White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al. Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1 Science 286 (5444), 1571-1577 (1999) 20036896	CDS	complement(3751..4860) /gene="DRA0208" /note="similar to PID:1001216 PID:1001269 percent identity: 66.97; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="peptide ABC transporter, permease protein" /protein_id="AAF12421.1" /db_xref="GI:6460716" /translation="MTGPEPTRPLQMAWRRYKSRPGVIAGWVLIYVIALISPEFL APYNTTQHEGYEYORPOTVHLIHQGRHLRPFVYVGGSTTKRDPVTFAKKTVPDTPAHLPL VQFPVKGPPSYFELGLOSRLHFLGVPDGDGEYGEKVPYFFFLGDDGLDFSRFLVG ACQSVTVGVGVLSIFALGILLGGVSGYFGWVDTLIQRVVEVLLSFLPLILLALAT LIPARMPSTWYTGIVAVLALIGWASLARVQGVLAARNLDYSAQAAMGASNLRLVI LRHTPLNSFLLVATIALPGYILGELSTLSFLGIGIKPEPMTSGLLLRDASKLEWS SYPMFLWPLGVFISVLAFFNFMGDALRDAADTQSR" complement(4857..5843) /gene="DRA0209" complement(4857..5843) /note="similar to PID:1001492 PID:1001545 percent identity: 62.93; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="peptide ABC transporter, permease protein" /protein_id="AAF12420.1" /db_xref="GI:6460715" /translation="MLCCMLNVTLRRLIGMLPTLLVLSVLCFLVLIKPGSGFIDQYL EDPFSKETVANITIGLDQPAWKQYLNMTGVVTRLDYGFSGAPVTSIIGERL GWTYIAALTALTLWLIAPVLGIYAFNRYGIPAQTANFLGVGLAIPDFIAALLLVV LVRLGGTVNGGLFRPMDIDAPWSPAKVLDLHLWIPVLAVGLEGVAGLMQRMRAST LDVLSQDITRTAKAGLPGQVRIWKHVRNAINPLISLAGLSLPTLSITLIISIVMS LPTIGLLYDLSLLNKDQYTMATLLMSLAFLLLLIGNLLSLDALAWADPRVRS" complement(5964..7804) /gene="DRA0210" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:AE000657 percent identity: 51.19; identified by sequence similarity; putative" 7764..8603 /gene="DRA0211" 7764..8603 /note="similar to GB:AL009126 percent identity: 60.78; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transcriptional regulator, GntR family" /protein_id="AAF12419.1" /db_xref="GI:6460714" /translation="MWSNNKTFSELVRKMYGCAVSSPAPLPPEPPAWVVALDPADAAP AYVQLAQGLRRRIESELGSLAGTAPAEALAAHGVSRVTLROGLLEAGLGLRRRR GSGTFVAGTPHTSRQLCLLSFSDVEYAGQGVPGARVLSFTQGRPTTPQALSGLTAA EQVYRLRLRTSDGPLVAEDVSTLPAALIGPLTAQDVTDASLYGLLQGRHLAPAIR HLRAQNAEALAHLLGVPLCAALLTTERVSWLASGRPEYARACRYGRDYDFVMDLRG EDH" 8603..9806 /gene="DRA0212" /note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to GB:I42023 SP:p44861 PID:1005712 PID:1220844 PID:1205001 percent identity: 56.04; identified by sequence similarity; putative" 9857..10774 /gene="DRA0213" 9857..10774 /note="similar to GB:U00096 SP:p76535 PID:1788768 percent identity: 63.89; identified by sequence similarity; putative" /codon_start=1
TITLE JOURNAL MEDLINE REFERENCE AUTHORS	1. 198677 /organism="Deinococcus radiodurans" /strain="R1" /db_xref="taxon:1299" /chromosome="2" complement(91..1926) /gene="DRA0206" complement(91..1926) /gene="DRA0206" /note="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="oligodeoxytase F, putative" /protein_id="AAF12423.1" /db_xref="GI:6460718" /translation="MLARACPLRAGPCRSRPAQSCPCMERERWARLPMTTQ SNLEAVEHLKMPDADQWASYAPEAALQEAOLTAGDVPAWLERNNLSARLOGSGS KLSTHADLHTDDDAIOQRYOTFVAELPQARASQALTEKGLAVPGVPGPDPALNR RFQDAALFPEANVDLGVTHNQINRHSVITGNQVOLGQOELTLPOAKOOMPSPDRA QREAAHALSASNAIAPELDAVMDLIRTRQLAHNADANFDRYRWKLEDVDTYP FOCLDFHSDVSWPLLSGLVGDIAERLGLSPWYNNRNLDOPEGDSLRPFAT GADLEALAOOTYELSDPLAARFQMRSGLLDRESRPGMKTHAYCNYPPTWPEPVL MNVGTAEDVRVLFHEVGHAFHGFYSGDROPLVNRSPTEFVEIPSMAMEFLTLDLH SHVFSDDLEARRKQLEGVIAELPWAQMDAFQHLIAXAEPEDVTIADLDAKLELD RTFPFFIDWSGLDESVRAGKWHYTHIFQVFFYIYEMCYLAAVAIVRAARQDPAAL GRYKDALRLGNTVSPPELYRAAGAEFRFDREYIRGLMGFLKEQLG" complement(2005..3645) /gene="DRA0207" complement(2005..3645) /note="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12422.1" /db_xref="GI:6460717" /translation="MNEFVLSPSLSSVAHALNRAKSLKEEMSKSRSLALLGALLL SACSPGNGQAPQAPQLPNTPEARPGTPSCVKRGLWYDAFGPGLKTPTEEVSTLVDTAR AMNVTLYVQGRGDCYCNNAAMPDDPPEVPAFGFDPLADVIKAKAKDIEVNAWII TTALWNSVTPPKSAEHAFTNHPGNAAGDNWLTVKADGTVKAGADWVLDPGHPDAAAY IRNNYSVKNYAVDGVTFMFRVRYPDYNPVGQVQWGYNETALARYAEAGTGPDP ADPQSWARKQOVTNLVRETAALAVKAPGVAATAATITYEGPADAAFAAASRPTAA VQDWLTWKSQYLDQNMNMYKRFVPLQAGFWDMNTFAAGLRQYPTHEQVSGAA IYLDIASTVSOIHKTOAAGLWAGYSYRTPKDKNVNEGKGAEEVLADLRLTLDAGG PFPFVYHAPLPAASVGLSGELPAARPGHSHVELVRAGGTWNVVAATSDGQGRYG FLNLPGQADVVRVDKILLPNVQATPGRVTHLEPITLN" complement(3751..4860) /gene="DRA0208"	gene	
TITLE JOURNAL FEATURES source	Submitted (02-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. 198677 /organism="Deinococcus radiodurans" /strain="R1" /db_xref="taxon:1299" /chromosome="2" complement(91..1926) /gene="DRA0206" complement(91..1926) /gene="DRA0206" /note="similar to GB:AE000520 percent identity: 43.92; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="oligodeoxytase F, putative" /protein_id="AAF12423.1" /db_xref="GI:6460718" /translation="MLARACPLRAGPCRSRPAQSCPCMERERWARLPMTTQ SNLEAVEHLKMPDADQWASYAPEAALQEAOLTAGDVPAWLERNNLSARLOGSGS KLSTHADLHTDDDAIOQRYOTFVAELPQARASQALTEKGLAVPGVPGPDPALNR RFQDAALFPEANVDLGVTHNQINRHSVITGNQVOLGQOELTLPOAKOOMPSPDRA QREAAHALSASNAIAPELDAVMDLIRTRQLAHNADANFDRYRWKLEDVDTYP FOCLDFHSDVSWPLLSGLVGDIAERLGLSPWYNNRNLDOPEGDSLRPFAT GADLEALAOOTYELSDPLAARFQMRSGLLDRESRPGMKTHAYCNYPPTWPEPVL MNVGTAEDVRVLFHEVGHAFHGFYSGDROPLVNRSPTEFVEIPSMAMEFLTLDLH SHVFSDDLEARRKQLEGVIAELPWAQMDAFQHLIAXAEPEDVTIADLDAKLELD RTFPFFIDWSGLDESVRAGKWHYTHIFQVFFYIYEMCYLAAVAIVRAARQDPAAL GRYKDALRLGNTVSPPELYRAAGAEFRFDREYIRGLMGFLKEQLG" complement(2005..3645) /gene="DRA0207" complement(2005..3645) /note="similar to PID:1653945 percent identity: 49.28; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12422.1" /db_xref="GI:6460717" /translation="MNEFVLSPSLSSVAHALNRAKSLKEEMSKSRSLALLGALLL SACSPGNGQAPQAPQLPNTPEARPGTPSCVKRGLWYDAFGPGLKTPTEEVSTLVDTAR AMNVTLYVQGRGDCYCNNAAMPDDPPEVPAFGFDPLADVIKAKAKDIEVNAWII TTALWNSVTPPKSAEHAFTNHPGNAAGDNWLTVKADGTVKAGADWVLDPGHPDAAAY IRNNYSVKNYAVDGVTFMFRVRYPDYNPVGQVQWGYNETALARYAEAGTGPDP ADPQSWARKQOVTNLVRETAALAVKAPGVAATAATITYEGPADAAFAAASRPTAA VQDWLTWKSQYLDQNMNMYKRFVPLQAGFWDMNTFAAGLRQYPTHEQVSGAA IYLDIASTVSOIHKTOAAGLWAGYSYRTPKDKNVNEGKGAEEVLADLRLTLDAGG PFPFVYHAPLPAASVGLSGELPAARPGHSHVELVRAGGTWNVVAATSDGQGRYG FLNLPGQADVVRVDKILLPNVQATPGRVTHLEPITLN" complement(3751..4860) /gene="DRA0208"	gene	

```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF12418.1"
/db_xref="GI:6460713"
/translation="MTDPRTEGVHPDYADLTLPADALIAALADUQLGAVRAVVAQA
POLTAALNAAPVQLERGVYVQAGTSGRLGLVDATETLTFSWPPERAVPLIAGGE
RAIQVQEGAPLLAAVDCQPVLLDTGPETISGSTRLKAGTAKTALNTLSSALMVRIGKL
VGLNLMVDVRAITNAKLEDRARLQVHATGADADAQAALSECQSVKTAIVMLKILGIGA
QEAAQLEGAGGHARGVQVGEALGTSAS"
gene
CDS
complement(10796..11398)
/gene="DRA0214"
complement(10796..11398)
/gene="DRA0214"
/notes="similar to GB:M99166 SP:P30849 PID:148264 GB:U00096
PID:1651499 percent identity: 54.36; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="trp repressor binding protein wrba, putative"
/protein_id="AAF12417.1"
/db_xref="GI:6460712"
/translation="MTAPVKLAIIFYSTGTGYMAQEAAGRAAGAEVRLIKVRFIT
APQDVIGDQARKANIEAMKDVPEATPADLEWAEAVFSSPTREGCATSQMRAFIIDL
GLWSSGKLANKTSMNTSAONVNGOETTLQTLMTAMHWGAVLTTPPGYTDVEVIFS
GNNPYGAVTANGPQLLENDRASIRHQVRQVELTAKLLG"
gene
CDS
11664..12059
/gene="DRA0215"
11664..12059
/gene="DRA0215"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF12499.1"
/db_xref="GI:6460794"
/translation="MGGLHPRQLHDLRLSDELMFGLNPREVAFLKGAGINAAKG
LTGMGNGTTPHPLAAGDPAFAALISDNEQARLGLGVRLVVAHHLTCLTLGAPLIDPND
TERVWORIEATRPPLFVDEVWAGREDEGA"
gene
CDS
12068..13069
/gene="DRA0216"
12068..13069
/gene="DRA0216"
/notes="similar to GB:U00096 PID:1788770 PID:1799860
PID:1799864 percent identity: 54.32; identified by
sequence similarity; putative"

alignment_scores:
Quality: 644.50 Length: 464
Ratio: 2.170 Gaps: 5
Percent Similarity: 64.009 Percent Identity: 30.603

alignment_block:
US-09-911-667A-2 x AE001863
..
Align seg 1/1 to: AE001863 from: 1 to: 198677

70 valgluaspAlaValcIuaIaIleGlyPheGlyLysPheGlnTrpIysLe 86
||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
77115 GTGACCGCGTCTGGCAGCACCTGGCGCTGGCCGCTTCAGTGAAGCT 77164

86 uSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetI 103
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
77165 GCTCGCCATCTCTGGCTGTACCTGGCGCGCGCAGCGATGAGGTGCTCC 77214

103 leLeuSerilleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSer 119
||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
77215 TGATGGCTTCGCGTGGCCGCATCAGTGGCGGCTTCGAGTCCGCAAA 77264

120 TrpGlnValAla.....LeuLeuThrSerValValPheValGlyMetMe 134
: : : ||| : : : : : : : : : : : : : : : : : : : : : : : :
77265 GGCTCGCGCGCGCGACGATGCTGTACCGGACTTTTGGCGGATGCT 77314

```


68653 GCTAATAGGAGCTTACACTTGGGTGTAGTTTCAGACAACACGGAAGGA 68702
149 149
68703 GCGAGTCCATCGTTTACTCCAATCACTGATTTTGTTCACACACAT 68752
149 149
68753 TATGCATAATAATTGATCTCTTGGTGATGCTTATTCCTTTAGCTTAGCA 68802
149 149
68803 ATGGGTCAAAAAGATGGTGTGAACCTGTTTCCATGCACAGCTTGTC 68852
149 149
68853 CAGTTGATAGTTTAGCCATAGTTATTAAAGACCGGACCGGCGGCTC 68902
149 149
68903 GGACCGGAAAAACAGAACTGGAGCCTCTGCCGGTTCCGTCCAGTTAAA 68952
149 149
68953 AGACCGGATATGCAATCAAGCCGTATAAACCGTTAGGCTGTCGGGTTT 69002
150 LysThrGlyLeuLysIle..... 155
69003 TCATGAAAACCGCTATAGAACCGGAGAAAAAACCATGGTTGAACCGGCC 69052
155 155
69053 TTGCGAGGAGTGGGTGCATCAAGGAGGAGCAGTCGGGGAAAAATGAGGGA 69102
155 155
69103 AAAAGAGCCCTCACCACCACCGCTATCATCTTGTGTGTTTGAAG 69152
155 155
69153 GAATGAGTAATATATTTAACTATTGAACCGGTTAGACAAATGGTTC 69202
155 155
69203 AACCGGTTGAACCACTGAACCATCAACCAAGAGCCTCACCGGTTCAATT 69252
155 155
69253 ACCGGTCCAGTCTCAATAACTATGTTTTAGCTAGTAACAACTGGC 69302
155 155
69303 AAAGGTCTTAGATAATTGATCAGTTAGACATTTCTATCTTACTTTCAGAA 69352
156 Ser.ValLeuTyr 159
69353 ATATCCTTGAACCACTCCCTCAAAATTTATGTTCTCGAAATCATATATTTT 69402
159 rp..... ThrLeuTyrTyr..... 163
69403 GGTGATGCTTATTTTCATGATATTGCAACACACTACCTTGTGATATATTCCTCT 69452
163 163
69453 TGGTGTGCAGCACAGAGCTCATACATTTCTGTGATTTGTAATTACAG 69502
164 GlyIleLeuS 167
69503 GAGAGGATTTCTTTTACCGCATCGTGACTAGTGGAGCTGGATTCCTCA 69552
167 erAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuVal 183
183 183
69553 GTGCTTTTGCACCTAACTATGTGTCACACTAATTTCTCTAAGATTTCTAGTT 69602

184 GlyPheGlyIleGlyGlyValProGlnSerValThrLeuTyrAlaGluPh 200
184 184
69603 GGTATTGGATTGGAGGAGCCTCTTGCATCTTGGTCTTGGAAAT 69652
200 eLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPhe 217
69653 TGTTCCTGCTCCAACTAGAGGAAGTGGATGGTAGTATTTTCTGCCTTT 69702
217 rpAlaIleGlyThrValPheGlu..... 224
69703 GGACTGTTGGACCATCTTCGAGGCTTCTCTGCTGGGTATGATATTTT 69752
224 224
69753 TCCCTGCTATTTTTAACTGGAATATCTGACTTTTACTTTTGGTGGTTAT 69802
224 224
69803 CCAAGGTCAATAGAAAGTGCCCTGTCTGCTCATTCATTCAATGAAGC 69852
225 Val 225
69853 CTACATCTCATTTGAAAAATGTATGTAACATATCTTTAGGCCCTGTCAAT 69902
226 ValLeuAlaValPhe..... ValMetProSerLe 235
69903 ATTGTAACATCTTTTCATCATGCTAATTGCAGACAGTAAATGCCCAAT 69952
235 uGlyTrpArgTrpLeuLeuIleLeuSerAlaValPro..... LeuL 249
69953 TGGCTGAGGTGGTGTGCTAGCATTTCTGCTGCCATCTTTCTGCTCTGC 70002
249 euLeuPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAsp 265
70003 TTTTATTTTATGTTATT..... ACACCAGAGTCAACCAAGGTTCTCTA 70043
266 ValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAl 282
70044 TGCATGAAGGCGAGTAACAGAACCCATGATGATGACTGGAGAAATGGC 70093
282 aThrGluAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArg. 298
70094 TAGATTGAACAATGTACAGCTCCCTCTGGAAGGCTTGATCTGACAAA 70143
298 298
70144 ACATTGAGCTAGATGAACCTTTCACATCTGAGTCTACAACACTTCTTGCT 70193
299 GlnGluAspArgGlyLysMetLargAsp..... 307
70194 GATCCTGAAGAAGATGACACCATCATTTGAACACAAAGTTTCCAGTTTGG 70243
308 LeuPheThrProHisPheArgTrpThrL 318
70244 AGGTTTGTCCGTCGGAAGCTGCTGCGCCAAATTTGATACGTGCAACTC 70293
318 euLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeu 334
70294 TGCCTGTGGATGACTTCTTTGGAAATGCTATTTCTATATTGGAAT 70343
335 ValLeuLeuThrThrGluPheGlnAlaGlyAspValCysGlyIleSe 351
70344 GTTCTGCTCACATCGGAGTTAAGTAACGGTAATAGGATATGTGA..... 70388
351 rSerArgLysLysAlaValGluAlaLysCysSer. LeuAlaCysGluTyr 367
70389 AAAGAGGAGGTGAATCATACATCAATGATGATCAAGCTTAT 70431
368 LeuSerGluGluAspTyrMetAspLeu..... LeuTrpThrThrLe 381
70432 ACAAAAACGTTTTTATATATCTAGCTTTGCGAGGTTTGTGTTTCTGTACATT 70481

TITLE	Direct Submission	Source
JOURNAL	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA	
FEATURES	Location/Qualifiers	
1..11286	/organism="Agrobacterium tumefaciens"	
	/strain="C58"	
	/db_xref="taxon:358"	
118..1128	/gene="AGR_C_4098"	gene
118..1128	/gene="AGR_C_4098"	CDS
	/note="(AP001509) BH0731"	
	unknown conserved protein"	
	/codon_start=1	
	/transl_table=11	
	/product="AGR_C_4098p"	
	/protein_id="AAK87995.1"	
	/db_xref="GI:15157409"	
	/translation="MGVMTGKTLDEKAIVASAEKALSDIAAIRTEVSKVIFGQPKV ONTLAILSGHALLVGPCLAKTLVTTLGTIVGLDANRIQFTPLDMLPDSILGSEVM DQDNGRRSFRFKGPIFGQLLMADEINRASPTOSALQAOEYHITMAGOTYELPK PFHVLTATNPLOEGYTPPEAGLDFLLQVDVYDPELAAEQIILLDTTGTAGGARP VIGAEIRMETQALIRQMPVSDKVVDAIILSLKRSARPGHGNKLTIDNVVWGPGRAGOS LMLTARALYEGRLAPSLDDVHALAEPVLEHRMALTFAAREGMSVRDVIIALVEQA KN"	
1159..2079	/gene="AGR_C_4100"	gene
1159..2079	/gene="AGR_C_4100"	CDS
	/note="hypothetical protein"	
	/codon_start=1	
	/transl_table=11	
	/product="AGR_C_4100p"	
	/protein_id="AAK87996.1"	
	/db_xref="GI:15157410"	
	/translation="MPVASIQIVDKTPGSEVLARARORAAALVPCDMVAKRIANTVT AGWRRKRGIGENWFQRPYAEGLSRISLIDWRSSRDHITYVRDRENEAAHTIHLWA DMSPMFKSTLGSVSKESALVLMALAEILARSGERICGPIGMEPEVPSARNAARLAA AAIIMHAPLTGMPDTSIMRGASDILVIGDFLDDAKNVMERISPLARRGLURGHVVEIAD PAELTFYSGRTFESDPTGEKLVSGRAETIRIDYTRAYLARREALSSSLRLRLGMWF FHRTHLASEALVAVHMYLSGAPAQGGHR"	
2076..4886	/gene="AGR_C_4102"	gene
2076..4886	/gene="AGR_C_4102"	CDS
	/note="hypothetical protein"	
	/codon_start=1	
	/transl_table=11	
	/product="AGR_C_4102p"	
	/protein_id="AAK87997.1"	
	/db_xref="GI:15157411"	
	/translation="MSALPFVFTSPYILFGLLALPAIWMLLRLTPPRPKAEVFPPLKI LATVLRKEETPSKSPWMLTILRLAAAVIFALADPVNPNRNSIAGSGPLVLVDONS WASADPWERRVATACALIGDAERADRPVSIPTADAEDAVSGTTTAVLADKLAAAPK PLVDPDRARTAEITLNGTAPGTLAYIADGVOTAOESODSOGRILANGTSFAPGAET GKAVAAITADTNADAMSVTLSRLDTAQASLTVNAQDSOGRILANGTSFAPGAET TATYAEPELRNDNFARUSIEGVSTAGAVHLLDGFRRRRVALLAGETGNDFOPLLOPL YIISRALQPFVDLIPQAOADLQAIPEILQSNFSVIVMADIGRLPQETIYVPLORWLGA GGTILIRFAGPRLAAAPADPLPVTLTQGERALGGALSWAEPQLADFPFAGAFAGM KADGILIRKQVLAEPITPLDAERTWASLADGTPLVTLTNRVAGRIVLHVSAEASWSD PIISGHFVDMRLRIVOLSKAGGASANAPATALLPFRLLTAEGMSAEIGTARPLEMR AGQPRTGFDNPPGLYCTEDGFVALNLLPAGATRLPLDTSAGGVSTNEALIGETAKS LRPLAIFAAFMMLADSLVLFMNGAFARLPKARSATVATLLALGAFAMSPTDARA DDPPKGDGEIFERLDTHLATYRTGEDDVRISSEQGLQSEFELTWRTLTLEPQVGL DIKDELSPYIITYYVSATAMPSSAAISRIDYMRAGGTVLFDTRDQFSSLDIGAT SANGERLQAILANIDIPLEVPEDHVLTRSFYLLTNFPGRYNSGLWSVRSRGAOKT TSGLSGGDGTPTMITGNDNFAGAWDAQGAPVLPVTPPDDEMREHAFERSGVNIMY MLTGNYKADQVHVHVALLERLQ"	
4888..6972	/gene="AGR_C_4105"	gene
4888..6972	/gene="AGR_C_4105"	CDS

```

/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_4105p"
/protein_id="AAK87998.1"
/db_xref="GI:15157412"
/translators="MTLTAPELPMVIAIALAVAAALASFGLWRGVKGAWLRLGLAV
ALLAAIANPLLTOEREPELSTVVPVLDRSQSDVQDPQWDTALET.LKDLRSRPR
LEPRIVRNGSDSPQSLFSAUSSAVSDVSFVSGGAIF.LSDGOI.HDI.PNALNA
FQALGFRPVHGLITGKADEFORRIEIVRAPRFGIVNEEQULTLRVFDGRPAGGSA
EVTVMKGEETIQAOTPGQPTPFKYPKRGCNMMVEFSAELPGEVTAANNRAVHLI
RGINLRNLVILISGEQHAGEARWRLNLLKSDASIVLHFTILRPPKQDGTPI.NELSLI
APPTRELVK.IINDFL.IFDYQHRGVLPILYDY.IAQYVEKGALL.IAAGPEHAGQ
DSIANTPLASVLPAGQPTGOVHQAAFPRLSEAGKKHPVTRGLESGNEFPQGWRFRT
ISVPPQGOCTVMLIGDQDPLVILNRQEGRVAMLLSDQHLWARGFEGGQPHVALYR
IAHLMKEPELEEAUTARANGFTLQITRQITGDDPGQASVFPFSGTKENNAFTTPE
GLYRIERMTEPELTKNGEDTTLVHGVADEPAFKAMITTELTPLADATKGSVH
RVANSGR.IALPOTPLVRGDIRVADEDRMLIRLTDYTLVKGNTLPLFAGFAGLAIL
LAFSALMWREG"
7017..7475
/genes="AGR_C_4106"
7017..7475
CDS
/genes="AGR_C_4106"
/note="(L25828) ORF 12"
/codon_start=1
/transl_table=11
/product="AGR_C_4106p"
/protein_id="AAK87999.1"
/db_xref="GI:15157413"
/translators="MKPKSGAVRRLEKNTMNFVLSDAEAELKAI.HDPLVAYNIARFE
GESDKRDLNITIRNDNSVTFGLVGHGTARGMLYVOLLFVPEAMRQGTAPKLLAMAEF
EARKGCMGAVIDTMNPDAURYERYCFTKIGSLGPLSSQSGSITWLEKRF"
complement(7485..8798)
/genes="AGR_C_4107"
complement(7485..8798)
/genes="AGR_C_4107"
/note="probable sugar transporter - Deinococcus
radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_4107p"
/protein_id="AAK88000.1"
/db_xref="GI:15157414"
/translators="MTAITMNDALERAGAYORRLMGIFGLWAAADAMQVLAVGTFVA
ASIAATFGLTVPQALQTGTAFFGCMILGAAGFCRLADRYGRRVLI.VTVACDALFVL
SIFSDFITLILVRLFTLGAAGVGTPLVDVAMMAEFLPAKNRGMVLV.FLEGFWAVGTLI
VALAAGSLAGVADANRYIFAVTAPGVLGLCLFLPESPLYLLRTGRSEAKAVVLI
NRLMVNVRSLADGTLQFDPDTKGGIFSPALRORSMTILAIWLVSVSYCYGVFTW
MPKILAGDGFVRYGLVLLVLAQIPGTAIAAGYVKNRKRPTLIGFCLLSALCGI
LFTLASAGAMTAALLINSALFTGALQIATYFPELTPESRATGMAAGAGAWARLGGI
LAPSLIYXIAQGFCAIGTIFAGLLVLAATAIMNAETRQALT"
complement(8888..9493)
/genes="AGR_C_4108"
complement(8888..9493)
/genes="AGR_C_4108"
/note="probable acetyltransferase VC0655 {imported} -
Vibrio cholerae (group O1 strain N16961)"
/codon_start=1
/transl_table=11
/product="AGR_C_4108p"
/protein_id="AAK88001.1"
/db_xref="GI:15157415"
/translators="MSHPHSGFGLSKHDLVYLTEDASHDAIIDLINEAFAFGPRH
TRAAIRREQDHPDRSLFSVCADDDGETIASRVTPVLAGVKGMLGLAVLRPSHKN
GIGRELIVRTAAAKRGESEAVILIGDPPYMLPGFEKVAIDFPDGPVDPNRLVLI
PVGDDVQHORLAKGIKWRKCEAAMPAADDESFEELRVYG"
9386..10746
/genes="AGR_C_4109"
9386..10746
CDS
/genes="AGR_C_4109"
/note="HYPOTHETICAL 37.4 KD PROTEIN IN EXUR-TDCC
INTERGENIC REGION (O328)"
/codon_start=1

```

```
166 LeuSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeu 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8495 CTTTCGATATTTACCCGAGTTTACCATTTCTGCTCGCTCGCTGCTTTCT 8446

182 uValGlyPheGlyLeuGly...ValProGlnSerValThrLeuTyrA 198
|:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
8445 CACCGAGCTCGGTGGGCGGCACCTTCCCGGTGGATATGCGATGATGG 8396

198 IaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGlu 214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8395 CGAATTCCTGCCTCCCAAAATCGCGGCGCTGCTGTTTCTCGAA 8346

215 ValPheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheVa 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8345 GGTTCCTCGCGGTGGACACACTGATC...GTGGCACTCGCGCTTGGG 8299

231 IMetProSerLeuGly...TyrArgTrpLeuLeuIleLeu 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8298 TGCCAGCTCTCGCGGCTGGCTGATGCTCGGCTATATTTTGGGTGA 8249

244 eAlaValProLeuLeuPheAlaValLeuCysPheTrpLeuProGlu 260
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8248 CGGCTTCCCGCGCTTCTTGGCTCGGCTTCTGCTTCTGCTGCGGAA 8199

261 SerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaTh 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
8198 TCGCCGCTCTACCTGCTCGGACGGGCGATCCGAAGAAGCAAGCGCT 8149

277 rLeuLysArgIleAlaThrGluAsnGly...AlaProMetProLeuGlyL 293
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8148 CGTCAACGCATGCTGGTGGTGAACGGCGCTGCACCGCTCGATGCGGAA 8099

293 ysLeuIleIleSerArgGlnGluAspArgGlyLysMetArgAspLeuPhe 309
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8098 CGGCGCTTTTC...CAGCTCAGCTGACCAAGCGCGCGGATATTT 8055

310 ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSe 326
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8054 TCGCGCGCTCGCGCAGCGAGCATGATCTCGGCATCTGCTTCT 8005

326 rAsnAlaPheSerTyrTyrGlyLeu...ValLeuLeuThrThrGluLeuP 342
:|:||||| :||||| :||| :||| :||| :||| :||| :||| :||| :|||
8004 CGTCTCCCTTCTTATATGCGCTTCCACCTGATGCGCGGCAACTG. 7956

342 heGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAlaValGlu 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7955 .....GCAGGAGCGGTTTCGGTTTC..... 7935

359 AlaLysCysSerLeuAlaCysGluTyrLeuSerGluAspTyrMetAs 375
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
7934 .....GTGCGCGGTACGGGT 7918

375 pleuLeuTrpThrThrLeuSerGluPheProGlyValLeuValThrLeuT 392
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
7917 TCTGTGTGTGGCTGGCGGCGAGATCCCGGTTATGCAATGCTGCT 7869

392 rPileIleAspArgLeuGlyArgLysLysThrMet...AlaLeuCysPhe 407
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
7867 ATGGCGTCGAGAGTGGGTGCGAAGCGGCGCTGATCGGTTCTGCGCTG 7818

408 VallIlePheSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAs 424
||| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7817 CTGTGGCGGCTCGCTGCTCTCTTACGCTCGCCTCGCGGTGCGAT 7768

424 nValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyP 441
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
7767 GATTCGCCCTCGCTTCTCATCATGAGCTTCGCGCTGCTC...GCCACT 7721

441 heGlnAlaTyrValTyrThrProGluValTyrProThrAlaThrArg 457
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
7720 GGGGTGCGCTTTATCGCTTTACCGCGGAATTTTACCCGACGCGGCGG 7671
```

```
458 AlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuI 474
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7670 GCGACGGCATGGGCTCGGGTGCCTATGCAAGGCTCGGGCGCTGT 7621

474 eThrProPheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrL 491
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7620 GCGCGCTCACGCTGCTGTTATGCCATTCGCGAGGTTTGGGTGCGGA 7571

491 euAlaValTyrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPhe 507
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
7570 TTGGCATCTTTCGCGGCTACTGCTGTCGCCCATTCGCGGACCATG 7521

508 LeuProIleGluThrLys 513
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
7520 ATCAATCGGAAACAGG 7503

seq_name: gb_ba:AE007266

seq_documentation_block:
LOCUS AE007266 11856 bp DNA BCT 15-AUG-2001
DEFINITION Sinorhizobium meliloti plasmid pSymA section 72 of 121 of the
complete plasmid sequence.
ACCESSION AE007266 AE006469
VERSION AE007266.1 GI:14523915
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group:
Rhizobiaceae; Sinorhizobium.
REFERENCE
1 (bases 1 to 11856)
Barnett,M.J., Fisher,R.F., Jones,T., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kelman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSymA megaplasmid
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
11481432
JOURNAL
PUBMED
2 (bases 1 to 11856)
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kelman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
Location/Qualifiers
1..11856
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
complement(116..1468)
/gene="Sma1447"
complement(116..1468)
/gene="Sma1447"
/notes="glimmer prediction; global similarity to BenK, a
benzoate transport protein in P. putida; Pfam domain
homology to sugar and other transporters"
/codon_start=1
/transl_table=11
/product="putative transmembrane transport protein"
/protein_id="AAK653916"
/db_xref="GI:14523916"
/translation="MLIAAROLEKGRMTTITVDDALDRAGTGTYORRLMAIFGLVNA
ADAMQVLAVGFTAASTATFGLTVPORLQGTGLFGLNLFCAAGFGLADRIGRRRL
ATVACDAGLLSVFAQDTVLLLRFLTGVAAVGGTLPVDYAMAEFLFARNRRL
VMLEGFMAVGTLLVLAANAASLAGADAWRYIFAVTAIPALIGVGLRFLVPESPLYL
LRIGKTSKAIIVDEILVNGKMLAGASLPPPTASAGIFSDLRKRLMLTALW
FLVSIYGVFTWMPRLAGEGFVGYGLVVLALAQIPGLAALAYGVKWRRT
LIGFCLLSALGCLLFVAAGTAMLIGVSLIMSALLGTWGLYAYTPELYPTASRATG
```

```

MGAGAMARLGLGLLAPSLMLGUVAGSFGLAIGFAGLIGFAGLLLVAAVAFLDAETRVRVSLA
/
complement(1918. .3108)
/gene="Sma1450"
CDS
complement(1918. .3108)
/gene="Sma1450"
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to PhaA, a
thiolase from P. putida (GI:1857942); global similarity
to acetyl-CoA acetyltransferase from C. acetobutylicum and
A. eutrophus (GI: 1174677); Pfam domain homology to
thiolase"
/codon_start=1
/transl_table=11
/product="Probable thiolase"
/protein_id="AAK65449.1"
/db_xref="GI:14523917"
/translation="MNKQDPVIVCGARTPLGSGFQGLKDLISAADLGAIAIVDALKRA
GLAPDAVEMFGCVLTACGGQAPARQAALGAGLPGVGATTVNMKCGSMKAAMLAH
DLIKAESASIVVAGMESMTNAPYLIDRARQGYRIGHQKVLMHFDLGLEDAYDKGR
MGSFAFDCAEATQFTFSAQDEYAIASLEKAQKASADGSFAEIEVPLSLASGKERTVN
LDQPOKARLDKIPLLKPAFRDGGITTAANASSISDGAALVLMRRSADKQGIPLA
VICGHATHADAPSLPTAPIGAIKALCRRIGWDIGEDVLFINEAFVAVPMAARELG
LDAENVHGGACALGHPIGASSARVITLVNALLRRGLRRGLASVCIGGGEATAVV
FISG"
complement(3133. .3900)
/gene="Sma1452"
CDS
complement(3133. .3900)
/gene="Sma1452"
/function="Small molecule metabolism; Central intermediary
metabolism"
/note="glimmer prediction; global similarity to eukaryotic
3-hydroxyacyl-CoA type II (GI: 2569966); Pfam domain
similarity to short chain"
/codon_start=1
/transl_table=11
/product="Probable"
/protein_id="AAK65450.1"
/db_xref="GI:14523918"
/translation="MOLKSRVFIIVTCASSGLGNAVTRMLAQRCATVGLDLDLKPAGEE
PAELGAAYFRNADVTNEADAFAAFKQEPGHVHVCAGTAPGKILGRSGPH
ALDSFARTVANLIGTFNIRLAAEYMSQGEDDAGDGERVINTASIAAFDGGIGQAA
YAAKGGVAALTLPAARELARFGIRVITAPGIDFPMWAGMPQVDQDALAASVPFPP
RLGRAEYAAVKHICENTMLNGEIVRLDGLARMAPR"
3995. .5065
/gene="Sma1454"
3995. .5065
/gene="Sma1454"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; global similarity to OruR from
P. aeruginosa (GI: 1657970), which regulates ornithine
utilization; Pfam domain similarity to N-terminal
helix-turn-helix motif"
/codon_start=1
/transl_table=11
/product="Putative transcription factor"
/protein_id="AAK65451.1"
/db_xref="GI:14523919"
/translation="MFNEKGPISHMERRMIAPCFMDDTLECLRRHGVDAQPLLAQAGLP
SIVTGPVSANQYCAFHVAQAQNDDEFFEGCAGPMRAGSFALLCHLTLATLEHALR
LPIRWDFRVCSPAGTDLVLENGLAQIVLKDGATRSFAYRTFWILHGVNCLIGRR
RAPANILVRYRHADGLSAAIQRLHATAPSAWPGFEATAARMRIAPATLRRRLRQEG
QTVRSIKEDLRALAMEALADGRTNVAQLAVELGFSFSAFRAFRKWTGSKSPAQRFR
NATOTDFAESSSPKRKTQPOES"
complement(5047. .5565)
/gene="Sma1455"
5565. .5565
/gene="Sma1455"
complement(5047. .5565)
/gene="Sma1455"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11

```

```

/product="Hypothetical protein"
/protein_id="AAK65452.1"
/db_xref="GI:14523920"
/translation="MKRTGTCYRAHDPRWAFKPSGDGAAIKGARFNPKGVKTYLAL
SMTAVKANQGFARHIDPCVLCSDIDCADIADLTAEGRAGHGVSLFEACMSWAGA
FAEGRRPASWAIYNRLHPRGTAGILVPSFAPGTEAGDRNLVLTWGPDLPHRLDVIDP
SQRLPKDOLSWG"
complement(5562. .5951)
/gene="Sma1456"
5951. .5951
/gene="Sma1456"
/function="Miscellaneous; Unknown"
/note="glimmer prediction"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK65453.1"
/db_xref="GI:14523921"
/translation="MAQAKIQSDGGLJLSYMDKGGKIAYOVAVGFQMSKTLQAE
AGLARETLVRSRGTONTLRMLEIISRTWDAGKEQAMARYRAQPLPAFGGR
TABALVKDKAAARVYLDHMLGGA"
complement(6059. .7063)
/gene="Sma1459"
7063. .7063
/gene="Sma1459"
/function="Small molecule metabolism"
/note="glimmer prediction; global similarity to
malate-dehydrogenase from M. fervidus (GI: 80950)"
/codon_start=1
/transl_table=11
/product="Putative"
/protein_id="AAK65454.1"
/db_xref="GI:14523922"
/translation="WTPVETSMEEIEAVCLEALTLHCAAAATARAVAGAITRAEDCN
RVGLYLPFICRHLAIGKVDGEAVPEVTGTVTVTDARSQFAHPAIAAGTPTALIDL
AKAGLAAMAVRNSYNCLAGHVRPLADAGLIGCVSNAPASVAPPGATRALFGTNP
LAFVPSKEGAPTIVVDQSMASVTKEMLILRRDGEAIPIGMAODNGQPTTDAATGL
EGSLPAGRGKGANIALLEVALAALTGSAALTEASAFNEEGGPPHVGQFLAIDPD
HFAAGHSEAMONLVASHDAAGVRLPGFGRKQPCVCDADLANKKAVILSKSKNRQKHG"
complement(7060. .8184)
/gene="Sma1461"
8184. .8184
/gene="Sma1461"
/EC_number="5.5.1.1"
/function="Small molecule metabolism; Central intermediary
metabolism; amino acids"
/note="glimmer prediction; global similarity to muconate
cycloisomerase in P. putida (GI: 77758); Pfam domain
similarity to mandelate racemase / muconate lactonizing
enzyme family"
/codon_start=1
/transl_table=11
/product="Putative muconate cycloisomerase"
/protein_id="AAK65455.1"
/db_xref="GI:14523923"
/translation="MKRISNVRVRLVPLKQPHWSYGRISFAVNLIEADDGTV
GIGCEVAPDQTGTAALILYKLAKHLVGHSDVAPLIARFHOEYELGHGANIKRAAQ
IFSGIDMMDLQGLAGLVPVHQLLGGAKRQAVGYFYGQETAEELARDAAVHGAQG
ERYVYLKVGSEKLDLEITAAVRGEQDARLRLDANEGWSVHDAINNCRKLEYDIEF
IEQPTKSVSPIMAHVREKVGIPVADQAAFTLYDVVEICRQRAADMICIGPREIGGI
QPMKAAVAEAGLKICIHSSFTTGTTCAEHGIGLAIPNLDGNGIMHQLVQEDIV
SSDPLPKNGLDAFRKPLGCPQLAEDLVAGEGGRYAASR"
complement(8311. .9201)
/gene="Sma1462"
9201. .9201
/gene="Sma1462"
/function="Cell processes; Transport of small molecules"
/note="glimmer prediction; global similarity to
choline-binding protein precursor from B. subtilis
(GI:7404430) and several putative ABC transporter
proteins; adjacent ORFs encode other putative members of
the choline transport system"

```

alignment_scores:
 Quality: 540.50 Length: 458
 Ratio: 1.917 Gaps: 9
 Percent Similarity: 61.572 Percent Identity: 30.131

alignment_block:
 US-09-911-667A-2 x AE007266/rev ..

Align seg 1/1 to reverse of: AE007266 from: 1 to: 11856

```

67 ThrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGly 83
1423 ACAATACAGTAGACGATCGCTCGATCGCGCGCGACGGAATACCA 1374
83 nTrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG 100
1373 GCGCAGACTGATGGCAGATATTCGGTCTGGTTGGCTGCCGATGCCATGC 1324
100 LuMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArg 116
1323 AGGTGCTGGCGTCCGCTTCACCGCGCATCGATCGCGCGCACTTTCGCG 1274
117 LeuProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMe 133
1273 CTCACAGTTCGCGCAGCAGCTTCAGACGGCGCAGCGCTTTCCTCGGAAT 1224
133 tMetSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgL 150
1223 GCTGTTCCGCGCGCGGGTTTCGCGAGGCTCGCGACCGGATCGCGCCGCC 1174
150 yThrGlyLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeu 166
1173 GCGCGCTTCGATCGCCACCGCTCGCTCGGAGCGCTCTCGGTCTGCTG 1124
167 SerAlaPheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeu 183
1123 TCGGTCTTCGCGCAGCAGCTTACCGCTTTTGGCTTCTGCTCGCGTTCTC 1074
183 lGlyPheGlyIleGlyGly...ValProGlnSerValThrLeuTrpAlaG 199
1073 GCGCGCAGCTGTCGCGCGGACACTCGCGTGCAGTATCGGATGCGCGG 1024
199 LuPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluVal 215
1023 AGTTCCTGCGCGCGCGGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 974
216 PheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheValMe 232
973 TTTGGGCTGTCGCGACGCTCAFC...GTGGGCTCGCGCGCTGCGCTGCG 927
232 tProSerLeuGly.....TrpArgTrpLeuLeuIleLeuSerA 245
926 GAGCCTCGCGCGGCTCGCGATGCTCGCGATACATCTTCGCGATGACGG 877
245 lValProLeuLeuLeuPheAlaValLeuCysPheTrpLeuProGluSer 261
876 CAATCCGCGCAGTGTATGCGCGGCTCGGTTCCTGCTGCGCGGATCG 827
262 AlaArgTrpAspValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLe 278
826 CCGCTTTATCTGCTCGCGCTGGGAAAGACGTCGCGAGGCGAAGCCATCGT 777
278 uLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeu 295
776 CGATGAGATCTCTGCTGCTCAACGCG...AAATGCGACTTGGCGCGCGG 730
295 leuLeuSerArgGlnGluAspArgGlyLysMetArgAspLeuPheTrpPro 311
729 CATCGCTGGTTCGCGCGCGCGCGCGCGAGTCCGGGTATTTTTCGCGG 680
312 HisPheArgTrpThrThrLeuLeuLeuLeuTrpPheIleTrpPheSerAsn 328

```

```

679 GATTTCGCGAAGCGCAGCTCTGATGATCCTGCCCATCTGGTTTCTGGTTTC 630
328 aPheSerTrpTrpGlyLeuValLeu.....LeuThrThrG 340
629 CATATCTATTACGGCGTCTTACCTGGATCGCTCCGAGGCTCGCAGGCG 580
340 LuLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAla 356
579 AAGGCTTC.....GGCTTCGTTGCGCGC..... 557
357 ValGluAlaLysCysSerLeuAlaCysGluTrpLeuSerGluGluAspTy 373
556 .....TA 555
373 rMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuValT 390
554 CGGCTTCCTCGTCTCGCTGGCATTGGCCAGATCCCGGCTATCGCGTCG 505
390 hrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAlaLeuCys 406
504 CTGCTACGCTGTCGAGAAATGGGCGCGACGCGCGACCTG...ATAGGC 458
407 PheValIlePheSerPheCysSerLeuLeuLeuPheIleCysValGlyAr 423
457 TTTCTGTTTGTCTGTCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 408
423 gAsnValLeuLeuLeuPheIleAlaArgAlaPhe...IleSerG 439
407 GGCATGCTGATGCGCGTCTGCTCTCATCATGAGTTCGCGCTGCTGCTG 358
439 lYgLYPheGlnAlaAlaValTrpThrProGluValTrpProThrAla 455
357 GCACCTGGGTCGCGCTTATGCTATACGCGGAACTCTATCCGACCGCA 308
456 ThrArgAlaLeuLeuGlyLeuThrCysSerGlyMetAlaArgValGlyAl 472
307 TCGCGGCGGACAGGATGCGCGCGCGCGCTATGCGCGCTGCGCGG 258
472 aLeuIleThrProPheIleAlaGlnValMetLeuGluSerSerValTrpL 489
257 CTTCGTCGCGCGCTGCTGATGCGGCTGCTGCTGCGCGAGATTTTCGCTC 208
489 euThrLeuAlaValTrpSerGlyCysCysLeuLeuAlaAlaLeuAlaSer 505
207 TCGCCATCGGATCTTTCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCG 158
506 CysPheLeuProIleLeuThrLys 513
157 TTTCTGATGCGCGCGCGAGACGCGC 134

```

seq_name: gb_ba:AE001804

seq_documentation_block:

LOCUS AE001804 17641 bp DNA BCT 02-JUN-1999
 DEFINITION Thermotoga maritima section 116 of 136 of the complete genome.
 ACCESSION AE001804 AE000512
 VERSION AE001804.1 GI:4982160

KEYWORDS
 SOURCE Thermotoga maritima.

ORGANISM
 Bacteria; Thermotogales; Thermotoga.

REFERENCE
 1 (bases 1 to 17641)

AUTHORS
 Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
 Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
 McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
 Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
 Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
 Eisen,J.A., Fraser,C.M. et.al.

TITLE
 Evidence for lateral gene transfer between Archaea and bacteria
 from genome sequence of Thermotoga maritima

JOURNAL
 Nature 399 (6734), 323-329 (1999)

MEDLINE
 99287316

REFERENCE
 2 (bases 1 to 17641)

AUTHORS	Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Uitterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE	Direct Submission
JOURNAL	Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1. .17641
organism	"Thermotoga maritima"
db_xref	"taxon:2336"
gene	146. .580
CDS	/gene="TM1590"
gene	146. .580
CDS	/gene="TM1590"
note	"similar to GB:AE000657 percent identity: 74.38; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"translation initiation factor IF-3"
protein_id	"AAD36657.1"
db_xref	"GI:4982161"
translation	"MTGVPTPKALEAREKGLDLVLVAPNPPVAKIMDYGKYKYQ LTKQENKKPKVQMKMFRKLIDEDHQTKVKHLRRELDGCHKVYVVMFIGREHM FAEKGEILLERVTKDTELATVESPMEGRDMMVVLKPKNS"
gene	595. .792
CDS	/gene="TM1591"
gene	595. .792
CDS	/gene="TM1591"
note	"similar to SP:P55874 PID:1770008 GB:AL009126 percent identity: 65.62; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"ribosomal protein L35"
protein_id	"AAD36658.1"
db_xref	"GI:4982162"
translation	"MPKVKTNRSAAKRFRTKNGKIMRNHAYRSHKTGKKRNALRAL RKDDVVSSADKNVRLRLGKK"
gene	806. .1162
CDS	/gene="TM1592"
gene	806. .1162
CDS	/gene="TM1592"
note	"similar to SP:P55873 PID:1770009 GB:AL009126 percent identity: 69.57; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"ribosomal protein L20"
protein_id	"AAD36659.1"
db_xref	"GI:4982163"
translation	"MRVKRAVHAKKKKKYKKAAGYRGALSHRYKIAKOMYVYKWKWY SYVGRKKRDMRKLMTIRITNIAARNEGLKYSLEIHGLKLAGVSNKKMLSELAVNDP EAFKEYVIAKEALAS"
gene	1173. .1916
CDS	/gene="TM1593"
gene	1173. .1916
CDS	/gene="TM1593"
note	"similar to GB:Pyro_h percent identity: 72.27; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"conserved hypothetical protein"
protein_id	"AAD36660.1"
db_xref	"GI:4982164"
translation	"MEKTKKIAYNSGKGVGKTTVAVNLAVALAAEGYQVGLLDLDLH GPNVQMLGSLPPEGEKIVPAKYGSLKVFSLAMIOEGAPVIMRGLKHAIEQL TRDVMGDLDLIDLPDGTGDEALSTFQIKPDVAIVSTPQKAVAGDDVRRAINFYK RLSCGILGLVNNMSLYVCPNCGEKIVYVFGKETAEKLAEEFGIPLIARIPMDPEVSLS DEGRPAVYKRGTVIEEEKKIVKVLSL"
gene	complement(1882. .5502)
complement	(1882. .5502)
gene	/gene="TM1594"
complement	(1882. .5502)
gene	/gene="TM1594"
note	"similar to GB:AE000657 percent identity: 60.77; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"conserved hypothetical protein, GGDEF domain"
protein_id	"AAD36661.1"
db_xref	"GI:4982165"
translation	"MMKVRVKFRLRETYMGDEFLEDDGEDQVLKIVRAPVDRSFFFN EVAKLRNLNPNVLPKIKISDGKFLFPYVYHNLALENISDQVAKLILKFLFLS KVGVTIPALGMDLIVNDGVFLIPALISNIPDDKGVFSPERSRATVEVCRFLKI HGIEPHLETRESFDSREIRIPYIHRKEELIKBDIEIAQKPFPIILITGEORVGTK IASVLVDGUREQGYMWHQITSLDRLRWIWDASDELLELLKLDGQKVVILIDDLFGS DLLSFLFEESGLTIITKILVTSTKAFQFHKVYRLSPFTVEETIRFLERAIGKISE DOVKLIHLSLKGSLPGYVELLKFNFKNLRENIIVGIFKLLRELDSPERLSVLGOK FTGTELVLEKITGDYHDTLMSAYDSGVTTTEEGYRFALREFMYYNKJSENKKK NLUHEKLFENLPDILAKHLSLEDPRLKFFYVLYRVKHFMDYKTRSLI EYLRLEE FFEKPYSTIESLKMKLIFRIDPLGVKEKFNHFRFKTLRNPVDVESILQKDSLYDIY NLVLSRPHIRAGKAKQOELLNTIKRELKFNSTRELYLKAHLLYDLFSFSDREA LKEMNIATSEGLDQVMYRGALGVSRKQARLFEDMPILSYTYFLEGLYTHKRDFOKAEV NLTSLLYEGKITNFLVQLERLRKQARLFEDMPILSYTYFLEGLYTHKRDFOKAEV FRTLELEKHGIERALRGLVNNYLFSGDVESAKLSLEKDEPEFDFRGNFLRLRL AKDSDSELLIKWLERLETPQFFNEEYAVVTERLAKLDPEFVEFLELELRNENSS NLTALVYTFYKYFYSALGENPKAKYLRRAIFVYNLIGLREVSVKPEVEKTOIEEK KAPYLLLOFLETEKEFSOMMEFASARLSEVIPYVESIRIERTTKVMEESTSPI TTPMERDLEISPFRTVMFSLDKMDHIVGVETNLECKDEKTAWELETLEOPGNILT TILREKLYRDRSMKPLDTGLVSRWYFMERLEEYAYKSSRYKSLPSLIIMCDADDFKIN DQFGHVAGDKTLGWLGRKMKSVLRKSDLVGRYGGEEFIIPALPGTSLEEKIVAEKLRK AVMEDPENTYHITLSFGVAEYKNGEDPFETIKRADEALYLAKILGKNSVVVTEKLSRR SS"
gene	complement(5450. .6451)
gene	/gene="TM1595"
complement	(5450. .6451)
gene	/gene="TM1595"
note	"similar to GB:AE000657 percent identity: 58.22; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"conserved hypothetical protein"
protein_id	"AAD36662.1"
db_xref	"GI:4982166"
translation	"MDEIVKVLQSHDRILVVGHIMPDCVSSVLSLTGLKGLKRGV KAADYKIPYVFEKPYIDKIEENPDFPELLVVVDASPDRIKFODLLDKVPVSVI DHRTNTNCGNMWVDPSPAATQMIPIRNKALGVEDSNLATNLNYLGATDYGFFRH SNADYVFEDAYKLVKMGADAHFVAKELLENKREOFKLF AEVLERLQLENGKIAYS YIDYTLRHNCITDEDSAGYVGLRSIRGVEAVLMEFPRGKIHVSMRSKDMFNVE VAFELGGGHPRAGAVTFEGKKIEEVIPRVINHLKKFKREGVESESEKIPEGDVLGG"
gene	complement(6452. .7249)
gene	/gene="TM1596"
complement	(6452. .7249)
gene	/gene="TM1596"
note	"similar to SP:P46354 PID:944944 PID:1303996 GB:AL009126 percent identity: 68.54; identified by sequence similarity; putative"
codon_start	=1
transl_table	=11
product	"purine nucleoside phosphorylase"
protein_id	"AAD36663.1"
db_xref	"GI:4982167"
translation	"MMKIEEARTFISERTNLSPDILILGSGFGPPIEKVEDPVIID YKDIPIHPQPTVEHSGKLVFGRISDKPVMIMAGRPHLYEGHDPAIVAFVPIYALVYG VKGVVTNAAAGINPEFKGEIILVRDIINFMFRNPLRGPNDEKIGRFPDMSVDDP EWARKQERLSLKEGVYIGLVGPSYETPAETIRVFEKLGADLVGMSTVPIAAKHGCL KVVVFCVTNNAAGIITHGRLSHEEVVTRTKMAQGKIEKALTAVEVE"
gene	7299. .8363
gene	/gene="TM1597"
7299. .8363	
gene	/gene="TM1597"
note	"similar to percent identity: 0.00; identified by sequence similarity; putative"
codon_start	=1


```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36664.1"
/db_xref="GI:4982168"
/transl_table="MYPRLLNKLKEIENARKVVMASRRGIIVGVTKVTLGDPRE
AETLRKAGICILGESRIINVLRLMKKAGIEGPFMLRLPMSELVEVDKHFDFIMVSDP
DVAKKVDLSREMKRVKILYIMIDVGLDREGVWFKAVEIAOHCANIVGIGTINFGC
YGIIPTRKFEILLDIKKELKXNHSFNIEIVSGGCTPALALENGEIEGINKLRIG
EATVIGRDITNNRVIDWLSONTFLIEAEVIEVKEKSPVPLGRGLDVFGRKVDVDRG
IRKRAICAGEQIDISGRGLIPVDKGVLEVLHASSDHLVDVDFGDKVKGVDFRFRMTY
SCLLKAMTSPFVEKVEPST"
gene      8396..8965
CDS       /gene="TM1598"
          8396..8965
          /gene="TM1598"
          /note="similar to SP:P34086 PID:1045628 PID:987648
          GB:U00096 PID:178926 percent identity: 61.67; identified
          by sequence similarity; putative"
          /codon_start=1
/transl_table=11
/product="RNA polymerase sigma-E factor"
/protein_id="AAD36665.1"
/db_xref="GI:4982169"
/transl_table="MNEKKLVEALKKEDWAYEVLYREFAPKIGSIKASFLNTDUDVD
VVDVFLRIKFKIGKFRGDSKSLTWIYIAVNVCKDYLOKTKRNEFLDQFQKDNKE
EIIYIQEAKENLVDELVSITAERVOALEKLSPEDKLLIKLRDIDGLSYEEIASVLO
KPVGTVKSRHLHYARKRLGNLLKEGTVER"
gene      8946..9308
CDS       /gene="TM1599"
          8946..9308
          /gene="TM1599"
          /note="similar to percent identity: 0.00; identified by
          sequence similarity; putative"
          /codon_start=1

alignment_scores:
  Quality: 527.50      Length: 523
  Ratio: 1.758        Gaps: 15
  Percent Similarity: 57.361 Percent Identity: 29.063

alignment_block:
US-09-911-667A-2 x AE001804 ..
Align seg 1/1 to: AE001804 from: 1 to: 17641
13 ValValLysPheArgArgThrGlyGluSerAlaArgSerGluAsp...As 28
12113 GTCTTGAATTTGAAATCTGATGGATGCGCAAGACACACCTCTCTT 12162
28 pThrAlaSerGlyGluHisGluValGlnIleGluGlyValHisValGlyL 45
12163 AACTCTCTCTGGA.....GGAGTTTCAT.....C 12185
45 euGluAlaValGluLeuAspGlyAlaAlaValProLysGluPheAla 61
12186 TTCACACATAGAGCCCGGATGAAGAACAAATGGAAAGAACATCATGAGA 12235
62 AsnProThrAspThrPheMetValGluAspAla..... 73
12236 GAGCTGAAGAAAAAGGTTTCTCATAGAGGAGGATAAAAGATGAGAAAT 12285
74 .....ValGluAlaIleGlyPheGlyLysPheGlnTrpLysLeuS 87
12286 AGACGAAATAGTTGAAATAACGTGGACAGAAAACTCAGAGAGGTTTC 12335
87 erValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetile 103
12336 TCATCTCTCATCTATCCCTGGATGTTTCGATGCTGCGGCTGTTATGCTT 12385
104 LeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTr 120
12386 CTTCTTTTGTGCTTCTTACGTGATATAAAGGAATGGAATCTCACCTCTAC 12435

```

```

120 pGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSerSerS 137
12436 TCAGGGAGCTACCATCAAGTGTCTACCTTCCTTGGAAATGCTTTTGGAG 12485
137 erThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLysThrGlyLeu 153
12486 CGCTTCTGTGGGTTTCGTGCTGACCTTCTTGAAGGAAAGTTTCCAAT 12535
154 LysIleSerValLeuTrpThrLeuTyrTrpGlyIleLeuSerAlaPheAl 170
12536 CTGTTCTTTTCATCGTCAAGCATCACCTTCTTCTGTCGGCTTTC 12585
170 aproValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGlyI 187
12586 CTCATCGTTCGAAACTCTTTAGTCTCAGAGGCTTTCGGGTTTGGGT 12635
187 LeGlyGlyValProGlnSerValThrLeuTyr...AlaGluPheLeuPro 202
12636 ATGGAGGCTTCATGCCCTTCGTTCAACGGGTATCTTGAGAAATTCACGAGT 12685
203 MetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpAlaIle 219
12686 ATCAGACTCAGGGGAAGGTATCTTGTGCTTCTGGAGTCGAGCTGGCGGT 12735
219 eGlyThrValPheGluValValLeuAlaValPheValMetProSerLeuG 236
12736 TGGGAGTATACTGATAGGACTCTTCGCAGTGAATGCTTCTCCCGAAC... 12781
236 lyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuPheAla 252
12782 ..TGGAGATGGGTCCTCTGGATCTTCTCCATCGGG...TACCTTTTGT 12826
253 ValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeuSerGl 269
12827 CCGTGTTCTTAGGATGCCGAAACCCAAAGTACGCGTTTTCGAGGG 12876
269 yAsnGlnGluLysAlaIleAlaThrLeu...LysArgIleAlaThrGluA 285
12877 TGGAAAGAGCGCTGGAGAGTCTCTGGGTAAAGAGTGGAGGAGGAGG 12926
285 snGlyAlaPro.....MetProLeuGlyLysLeuIleIleSer 297
12927 TGGAGCTTCCAAAAAAGAGAAAGTTCCCATCTCTGCTCTTTG..... 12970
298 ArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHisPhear 314
12971 .....AAGAGGGAACATTTGAAAGAC..... 12991
314 gTrpThrThrLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerT 331
12992 .....ACAGTTGTGATCTGGATACGCTGGTTCGTTGTGAGCTTCGTCT 13034
331 yTrpGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVal 347
13035 ATTACGCTCTGTTCACCTGGCTCCGAGGATCTCTCTCTCG..... 13075
348 CysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuAl 364
13076 CTGGGTGTGAGTGTGTCAAATCTTC..... 13102
364 acysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThr... 380
13103 .....TGGTTCACAT 13112
381 .....LeuSerGluPheProGlyValLeuValThrLeuTrpIle 393
13113 TCATCATGATGGTGGCAGCTTCCCGGGTACCTTTCCGGCAGCTACTTC 13162
394 IleAspArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePh 410
13163 ATAGAAAAATGGGGAAGAAAGGCTTCTTGTGTTTACTTTCATAGGAAC 13212
410 eSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAsnValLeu 427

```

seq_name	gb_in	CEZK637
seq_documentation_block	40699 bp	DNA
LOCUS	CEZK637	INV
DEFINITION	Caenorhabditis elegans cosmid ZK637, complete sequence.	
ACCESSION	Z11115	
VERSION	Z11115.1	GI:6953
KEYWORDS	HTG.	
SOURCE	Caenorhabditis elegans.	
ORGANISM	Caenorhabditis elegans	
REFERENCE	1 (bases 1 to 40699)	
AUTHORS	none.	
TITLE	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium	
JOURNAL	Science 292 (5396), 2012-2018 (1998)	
MEDLINE	99069613	
REFERENCE	2 (bases 1 to 40699)	
AUTHORS	Craxton, M.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-MAY-1991) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or w@nematode.wustl.edu	
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information. Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is not the entire insert of clone ZK637. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions. The start of this sequence (1..180) overlaps with the end of sequence 222175. The end of this sequence (40696..40699) overlaps with the start of sequence Z11126. For a graphical representation of this sequence and its analysis see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=ZK637 .	

```
/gene="ZK637.4"
/notes="predicted using Genefinder"
/codon_start=1
/protein_id="CAA77451.1"
/db_xref="GI:3881666"
/db_xref="SWISS-PROT:P30637"
/translation="MKSNPXYFLMNDVQRQSKYSPKYPVNSLKERIKLEFLDYVIAPL
KLYLLSYPPDCLWNRKLRKASGVQVTPSPSEPHIDRLIHISOKQPS"
join(11124..11557,11606..11864,12073..12282,12411..12536)
/gene="ZK637.5"
join(11124..11557,11606..11864,12073..12282,12411..12536)
/gene="ZK637.5"
/notes="predicted using Genefinder"
Ar5A homologue
cdNA EST yk62f10.5 comes from this gene
cdNA EST yk62f10.3 comes from this gene
cdNA EST yk140f3.3 comes from this gene
cdNA EST yk176c11.3 comes from this gene
cdNA EST yk140f3.5 comes from this gene
cdNA EST yk176c11.5 comes from this gene
cdNA EST yk221f9.3 comes from this gene
cdNA EST yk221f9.5 comes from this gene
cdNA EST yk267c10.5 comes from this gene
cdNA EST yk507e12.3 comes from this gene
cdNA EST yk191h10.5 comes from this gene
cdNA EST yk507e12.5 comes from this gene"
/codon_start=1
/protein_id="CAA77452.1"
/db_xref="GI:3881667"
/db_xref="SWISS-PROT:P30632"
/translation="MSDQLEASIKNILEQTLKWIIVGGKGVGKTKTCSLSLAQLSK
VRERVLITSDPAHNTDAFESQRTPTTLVEGKFLFAMEIDSNPNGEVEMCNIEE
MLQNAQSGSGFGKDFQSFAGLPGIDEAFSGEMIKLIDSLDFVVFDTA
PTGHLRLLOFPPTLLEKFTKILSLQGMFGPMNFGMFGMGGSMNEMIEKMTITL
ESVKNNQAQKDPNCTTFCVCIAEFLSYETERLQELSKQGITDTHNIIVNOLLPPD
TDANGTVSCRKASQALQSKYLTIDILEYDFHVVKLPILLEAEVRGPGAILQFSERM
VDEPANKN"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19653,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7b"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19647,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7a"
complement(join(16630..16827,17583..17732,17779..17883,
17962..18372,18848..19006,19051..19647,19694..19826,
19876..19994,20104..20160))
/gene="ZK637.7a"
/notes="predicted using Genefinder"
cdNA EST yk34b2.3 comes from this gene
cdNA EST yk24e1.3 comes from this gene
cdNA EST yk12h1.3 comes from this gene
cdNA EST yk24e1.5 comes from this gene
cdNA EST yk12h1.5 comes from this gene
cdNA EST yk34b2.5 comes from this gene
cdNA EST yk353h6.3 comes from this gene
cdNA EST yk353h6.5 comes from this gene
cdNA EST yk270e5.3 comes from this gene
cdNA EST yk270e5.5 comes from this gene
cdNA EST yk224d7.3 comes from this gene
cdNA EST yk224d7.5 comes from this gene
cdNA EST yk212f5.3 comes from this gene
cdNA EST yk212f5.5 comes from this gene
cdNA EST yk220e8.3 comes from this gene
cdNA EST yk220e8.5 comes from this gene
cdNA EST yk207e10.5 comes from this gene
cdNA EST yk296g5.5 comes from this gene
cdNA EST yk350b7.5 comes from this gene
cdNA EST yk430g11.5 comes from this gene
cdNA EST yk561d3.3 comes from this gene"
/codon_start=1
```

```
/protein_id="CAA77454.2"
/db_xref="GI:14530698"
/db_xref="SWISS-PROT:P30630"
/translation="MSSAVSRPKKKAASDTSQDRTSSPYSLRSTSKVPSRYRNEELY
LSPRSIKRTGSPKSPAKRLNGRSDPSVNSLTRNSLTLMLAKAALDYESSSCALEY
IPKERRRPRLALUSPPAPSPNDLLAKDELEHONLVAGLDDLDNPNMTNEAVE
HRDQTFNNFSTDQERSAMMKOFKTYKNQTSQEDVSTFMRANTKLYNLLRYKKRQW
VMCEFFYSIDEQIFKEENEFATITRESFNLKNWNI.TRIEWSIRKLLGKPRRCSKY
FTEERMYLEEKMKIRSVVEGSLNDPSIDLDLPAKLPRPMVGNRVFARIRNPLPD
GIYSGIIDAIVPKGPRIFDKPDIPPTLVSDTEILDGKLDLLSIAFYIEQAASKLPS
GVPPFAAAYRDSHPLVRDVLVSRKIERSGGFLMGPNDRNLNGKNAEMVGNFPLKFL
```

```
alignment_scores:
  Quality: 511.50      Length: 458
  Ratio: 2.483        Gaps: 13
  Percent Similarity: 44.978  Percent Identity: 31.659

alignment_block:
US-09-911-667A-2 x CEZK637/rev ..

Align seg 1/1 to reverse of: CEZK637 from: 1 to: 40699

42 HisValGly.....LeuGluAlaValGluLeuAspGlyAl 54
|||||.....:|||||
1367 CATGTCGGAGATGATTTTCCAGCTTCGATCAACCTCGATGATCGTAC 1318
|||||.....:|||||

54 aAlaValProLysGluPheAlaAsnProThrAsp.....66
:|||||.....:|||||
1317 T.....GAACCTGGAGAACCTACAGATCAGATCAGATCAGATGATA 1280

67 .....ThrPheMetValGluAspAlaValGluAlaIleGlyPheGly 80
|||||.....:|||||
1279 GTGAAAAGACGTTTACAGTTGATGAGCTGTTGAAGCATTTGGATTTGGA 1230
|||||.....:|||||

81 LysPheGlnTrpLysLeuSerValLeuThrGlyLeuAlaTrp.....94
:|||||.....:|||||
1229 AGATTTACGCTCAAACTGCAATTTTGACAGGAATGGCATGGCTGATTT 1180

94 .....94

1179 CGAGGAGTGTTCAAAAAGAGTGTTCATCCAAATGAGTTTATATATAC 1130

94 .....94

1129 AATATATTTTCACAATTTGAATTTGAAAATTTGGAATTTTACCAAAAT 1080

94 .....94

1079 TAGATTGAAACTTTTGAAAATAAAAAAACATTTTAGAGCAGATTTCTG 1030

94 .....94

1029 TTGCATGACTGTAATGAAAACATTCAAAATTTATTTTAAAAATATCAAA 980

94 .....94

979 AAAATAGTCACGGAAGAACCCAGAAAATTTTAAACACCTCAGAAAAAAT 930

94 .....94

929 GAATCTTCAATATGATTTTAAAAAAAATTTTATTAACACACCTTTCCGCA 880

95 .....Met..AlaAspAlaMetGluMetIleLeuSerIleLeuAl 108
|||||.....:|||||
879 GAATTTTCAGATGCCCGCCCAATGGAATGATGTTTATTTATTCATTGATTC 830
|||||.....:|||||

108 aProGlnLeuHisCysGluTrpArgLeuProSerTrpGln.Val.....122
|||||.....:|||||
829 TCCAGCTCTGGCATGTGAATGGGAATATATCGTCAGTGCACACAGGTAAGTT 780
|||||.....:|||||

123 .....AlaLeuLeuThrSerVa 128
|||||.....:|||||

779 TCAATCAACGCAAAAAAAAATTTGTTGAGGCTCTTGTGACGACGCTG 730
```

128 lValPheValGlyMetMetSerSerThrLeuTrpGlyAsnIleSerA 145
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
729 CGTGTTCAGTGGAAATCATGTTCTCAAGTACATATTTTGGGCAAAATATG 680
145 spGlnTyrGlyArgLys
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
679 ATCGATTGGTCGACGAAAGATAGTTGAAATATAAATGTATATAATTCGA 630
151ThrGlyLeuLysIleSerValLeuTrpThrLeuTyr 162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
629 AATCTGTAAAAAATTTCCAGGACTAAACATATTTTCAACACTAGTTGCTGTGA 580
162 rTyr.GlyIleLeuSerAlaPheAlaProValTyrSerTrpIleLeuVal 178
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
579 TCATGGGTGTCATTCTCGAAATGTCACATTTTACGTTCTTCTCTATT 530
179 LeuArgGlyLeuValGlyPheGlyIleGlyGlyValProGln..... 192
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
529 TTCGTGGACTTACTGGATTGTTATCGGAGGTGTACCCCACTCGTTAGT 480
193S 193
479 TTTTAAATCAGAATATTTCCAAATCCCTATTTTAAAAAACTTCATCTTTC 430
193 er.ValThrLeuTyrAlaGluPheLeuProMetLysAlaArgAlaLysCys 209
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
429 AGTGTCACTTTATACGCTGAGTTTCTTCCAACTGCTCAACGTCGCAAAATG 380
209 sIleLeuIleGluValPheTrpAlaIleGlyThrValPheGluVal 226
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
379 TGTAGTTTGTATAGAAATCATTTTGGGCTATCGGAGCAGTTTTCGAAAGCTC 330
226 alLeuAlaValPheValMetProSerLeuGlyTrpArgTrpLeuLeuIle 242
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
329 TTCGTGCTATTTGTGATGGAATCATTTGGTTGGAGACTCTTATGTTTC 280
243 LeuSerAlaValProLeuLeuLeuPheAlaVal.LeuCysPhe..... 256
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
279 CTTCCTTCACCTACCTCTTGGAATTTTTCAGTGCCTCTCTTTGTAAGTTT 230
256 256
229 TATCAAAAAATTCGAGTTAGAAATTTCAAAATGTCGAATTTGAACCTAG 180
257TrpLeuPr 259
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
179 TATGCAAGTATAGTATACAACTTCAAAATGTTTCAGTGGCTCCC 130
259 oGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleA 276
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
129 AGAATCTGCTCGTTTGACATGGCTTCTGGACATCCCAACAGAGCTCTAG 80
276 laThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGly 292
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
79 AAACACTTCAAGCAGCAGCAGGAATGAATAGAGTACAGTTTGCCAACTGGA 30
293 LysLeuIleIleSer 297
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
29 AGATTAGTGTCTTCA 15
seq_name: qb_pat:AX127153

```

Corynebacterium.
1 (bases 1 to 309400)
Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
AUTHORS Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 7069 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
    source          Location/Qualifiers
                1..309400
                /organism="Corynebacterium glutamicum"
                /db_xref="taxon:1718"
                /note="Seq 1 to long (3.309,400) split in 11, seq 7069"
BASE COUNT      70133 a 86477 c 83115 g 59675 t
ORIGIN
alignment_scores
    Quality: 498.50      Length: 480
    Ratio: 1.701         Gaps: 14
    Percent Similarity: 61.042      Percent Identity: 27.708
alignment_block
US-09-911-667A-2 x AX127153 ..
Align seg 1/1 to: AX127153 from: 1 to: 309400
90 ThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetIleLeuSerIl 106
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179148 TCAGGTATCGGCTGGGCATTCGATGCCATCGACTGATCTCTTT 179197
106 eLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnValA 123
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179198 CGTCATGGCTGGCTGGCCACTCATTCGGGCGCTATCCCTACTCAAACTT 179247
123 lAlaLeuLhrSerValValPheValGlyMetMetSerSerSerThrLeu 139
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179248 CCCTGCTCGATCCATCGGTTTCGTTGGCATGGCATGGCGCTTCGCTG 179297
140 TrpGlyAsnIleSerAspGlnTrpGlyArgLysThrGlyLeuLysIleSe 156
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179298 GCGGTTTCTGCTGGCGGCAAGTTGGGGCGTcGa.....CAAGTTTT 179338
156 rValLeuTrpThrLeuTyrTyrGlyIle.....LeuSerAlaPheA 170
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179339 TGGCTTGTCTTTCGTAGTTATGGGTTTGCACCTGGCGCGTGGCGGCTT 179388
170 lAProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGly 186
:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179389 CAGTGTGCTGGCAATGTATGATGGGCGTGGCGTTCGTCGTTGGACTTGGC 179438
187 lIeGlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuPr 202
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179439 CTGGGCGCTCAACTCCCGTTGCACCTCACTCTGATTTCCGAGTTTCTCC 179488
202 oMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpAlaI 219
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179489 ACCAAAAGTTCTGGCGCATGCTGTTATCCTGAGGCGGTTTGGGCGC 179538
219 lGlyThrValPheGluValValLeuAlaValPheValMetProSerLeu 235
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179539 TGGCTGGGATCATGGCTGCAATCGTCGGAACTTTTTCGTGGCAGGATCC 179588
236 .....GlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLe 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179589 GATAATAGTTGGGCTTGGGCGTTAGCTCTTGCTGTGCTCCCTCAATTA 179638
250 uPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValL 267
:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179639 CGCGGTGTATGCTCGGACTCTCCGAATCATGACTGTTCTCTGGAGA 179688
267 euSerGlyAsnGlnGluLysAla.....lIeAlaThrLeuLysArg 280
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


OM of: US-09-911-667A-2 to: N_Geneseq_1101.* out_format : pfs

Date: Mar 8, 2002 3:37 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame; p2n_model -DEV=xlip
-Q/cdn2_1/USPTO_spo1/US09911667/runat_07032002_130319_29572/app_query.fasta_1.615
-DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOCPCL=0.000 -LOOPEXT=0.000
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09911667_@CGN1_1_239 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-911-667A-2
Query length: 548
Database: N_Geneseq_1101.*
Database sequences: 930621
Database length: 42862619
Search time (sec): 180.210000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ49686		2846.00	5399.21	1.2e-292	2562
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ49686		876.00	1638.21	3.6e-83	1906
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ49689		597.00	1123.63	1.6e-54	401
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68534		498.50	856.21	1.3e-39	309400
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68260		493.50	911.44	1.1e-42	1344
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH67999		493.50	910.62	1.2e-42	1440
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ42410		471.00	880.71	5.6e-41	480
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ44679		446.50	817.50	1.9e-37	1901
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:AAH08702		428.00	782.27	1.7e-35	1882
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AAV49559		424.00	776.11	3.7e-35	1662
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AAV49558		424.00	774.59	4.5e-35	1888
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH57556		413.00	743.58	2.4e-33	4366
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH57556		413.00	740.71	3.5e-33	2121
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH16697		407.00	734.48	7.8e-33	3580
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:AAV12391		401.00	729.36	1.5e-32	2102
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH11146		401.00	728.93	1.6e-32	2179
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH11147		398.00	723.13	3.3e-32	2191
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH83979		392.50	715.66	8.7e-32	1697
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:AAV12393		390.00	708.15	2.3e-31	2135
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH26353		389.50	706.48	2.8e-31	2268
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH26014		386.00	700.64	6.0e-31	2112
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH58819		384.00	696.73	9.9e-31	2127
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAV79585		384.00	696.49	1.0e-30	2171
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH62004		379.00	689.03	2.7e-30	1821
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH85820		377.00	681.42	7.0e-30	2501
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH85821		375.00	680.57	7.8e-30	1950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ29300		374.50	678.60	1.0e-29	2123
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH5034		374.00	675.73	1.5e-29	2493
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH5035		367.50	663.27	7.2e-29	2502
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAV79584		361.50	652.83	2.7e-28	2284
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH26880		357.50	647.88	5.2e-28	1831
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH09889		357.50	647.88	5.2e-28	1831
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH88044		348.00	631.05	4.5e-27	1638
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:AAH26902		332.00	598.78	2.8e-25	1888
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH09890		332.00	598.78	2.8e-25	1888
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH88053		329.50	591.67	7.0e-25	2297
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH66162		316.50	573.67	7.1e-24	1293
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH68527		316.50	506.96	3.7e-20	349980
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH67950		311.50	565.04	2.1e-23	1197
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAH95401		311.50	561.65	3.3e-23	1590

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ49686

seq_documentation_block:

ID AAZ49686 standard; cDNA; 2562 BP.

AC AAZ49686;

DT 07-APR-2000 (first entry)

DE Human organic cation transporter-like protein (OCT1p) cDNA.

KW Human; organic cation transporter-like protein; OCT1p; transporter;
transmembrane; norepinephrine; neuroprotective; neuroleptic; anticonvulsant;
antiParkinsonian; antidepressant; cellular process; cell proliferation;
screen; treatment; prevention; diagnosis; neurodegenerative disorder;
Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
CNS disorder; central nervous system; schizophrenia; depression;
behavioural; sleep disorder; eating disorder; Alzheimer's; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 185..1831

FT /*tag= a

FT /product= "OCT1p"

XX WO200000633-A1.

XX 06-JAN-2000.

XX 29-JUN-1999; 99WO-US14880.

XX 30-JUN-1998; 98US-0107932.

XX (MILL-) MILLENNIUM PHARM INC.

XX Goodearl AJ, Glucksmann MA;

XX WPI: 2000-137069/12.

XX P-PSDB: AAY44633.

XX New nucleic acid encoding human organic cation transporter-like

XX protein, used for prevention, treatment and diagnosis of e.g.

XX neurological, behavioural or sleep disorders

XX Claim 2; Fig 1; 100pp; English.

XX The present sequence is a cDNA encoding human OCT1p (organic cation transporter-like protein), a member of the superfamily of sugar and other transporter molecules that have 12 transmembrane domains. The sequence is derived from a human foetal brain cDNA library. The protein is highly expressed in brain tissue and has norepinephrine, neuroprotective, neuroleptic, anticonvulsant, antiparkinsonian, antidepressant activities. The present sequence is used to regulate a variety of cellular processes e.g. cell proliferation, differentiation and survival, screen OCT1p modulators and detect mutation in OCT1p gene. OCT1p modulators can be used to treat or prevent chronic neurodegenerative disorders (e.g. Alzheimer's, Parkinson's, Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic, depression), behavioural, sleep and eating disorders.

XX Sequence 2562 BP; 520 A; 682 C; 749 G; 609 T; 2 other;

alignment_scores:

Quality: 2846.00

Length: 548

Ratio: 5.193

Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-911-667A-2 x AA249686 ..

Align seg 1/1 to: AA249686 from: 1 to: 2562

```

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
|||||
185 ATGGAGGAGGACTTATTCAGCTAGGACAGCTGCCGGTTGTGAATTCG 234
|||||
17 gArqThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
|||||
235 TCGCACAGCGAGTGCAGGTGCAAGGTGCAGAGGACGACGCGCTTCAGGAGC 284
|||||
34 isGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
|||||
285 ATGAAGTCCAGATGAAGGGGTCCAGCTGGGCTAGAGCTGTGGAGCTG 334
|||||
51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
|||||
335 GATGATGGGACGCTGTGCCAAGGAGTTTGCCAAATCCACCGATGATAC 384
|||||
67 rPheMetValGluAspAlaValAlaIleGlyPheGlyLysPheGlnT 84
|||||
385 TTTTCATGGTGAAGATGCAGTGGAGCCATTTGGAAATTTTCAGT 434
|||||
84 rPlysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||
435 GGAAGCTGTCTGTTCTACCTGGCTTGGCTTGGATGGCTGATGCCATGGAG 484
|||||
101 MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
|||||
485 ATGATGATCTCAGCATCTCCAGCACACAGCTGCATTTGGCGAGTGGAGGCT 534
|||||
117 uProSerTrpGlnValAlaLeuLeuThrSerValValPheValIGlyMetM 134
|||||
535 CCCAAGCTGGCAGGTGCATTGTGACCTCGGTGGTCTTTGTAGGCATGA 584
|||||
134 etSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLys 150
|||||
585 TGTCCAGCTCCACGCTCTGGGGAAATATCTCAGACCAGTACGGCAGGAAA 634
|||||
151 ThrGlyLeuLysIleSerValLeuTrpThrLeuTrpTrpGlyIleLeuSe 167
|||||
635 ACAGGCTGAAGATCAGCGTGTCTGGACTGTACTATGTCATCTCTAG 684
|||||
167 rAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValG 184
|||||
685 TGCATTTGCGCCCGTGTATAGCTGGATCTCCTGGTGTCTCCGGGGCTGGTGG 734
|||||
184 lYPheGlyIleGlyValProGlnSerValThrLeuTyrAlaGluPhe 200
|||||
735 GCTTCGGGATCGAGGAGTTCGCCAGTCGGTGGAGCTGTATGCCGAGTTC 784
|||||
201 LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
|||||
785 CTTCCCATGAAGCCAGAGCTAAATGATTTGCTGATTGAGGTATTCTG 834
|||||
217 pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetProS 234
|||||
835 GCCCATCGGACAGTGTTCGAGGTGCTTGGCTGGCTGTCTGCTGATGCCA 884
|||||
234 erLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
|||||
885 GCCTGGGTGGCGTGGCTGCTATCTCTCAGCTGTCCGCTCCTCCTC 934
|||||
251 PheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLe 267
|||||
935 TTTGGCGTCTGCTTTCTGGCTGCCGAAAGTCAAGGTATGATGTGCT 984
|||||
267 uSerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284
|||||

```

```

985 GTCAGGGAAACAGGAAAGCAATCGCCACCTTAAAGAGGATAGCAACTG 1034
|||||
284 luAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 300
|||||
1035 AAAACGGAGCTCCCATCGCGCTGGGAAATCATCATCTCCAGACAGGAA 1084
|||||
301 AspArgGlyLysMetArgAspLeuPheThrProHisPheArgTrpThrTh 317
|||||
1085 CACCGAGCAAAATGAGGACCTTTTTCACACCCCATTTTAGATGGACAA 1134
|||||
317 rLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyL 334
|||||
1135 TTTGCTGCTGTGGTTATATGTTTTCATGATCTCTCTTACTACGGGT 1184
|||||
334 euValLeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIle 350
|||||
1185 TAGTTCTACTCACCACAGAACTCTTCCAGGAGGAGATGTCTGCGGCATC 1234
|||||
351 SerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTy 367
|||||
1235 TCCAGTCGGAAGAGCTGTAGAGGCAAAATGCACGCTGGCTGGCGACTA 1284
|||||
367 rLeuSerGluGluAspTyrMetAspLeuLeuTrpThrThrLeuSerGluP 384
|||||
1285 CCTGAGTGAGGAGGATTACATGGACTTGTGTGGACACCCCTCTCTGAGT 1334
|||||
384 reProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLys 400
|||||
1335 TTTCAGGTGCTCTGTGTGACTCTGTGGATTATTGACCGCTGGGGCGCAAG 1384
|||||
401 LysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLe 417
|||||
1385 AAGACCATGGCCCTGTGCTTTGTCTCATCTTCTCTCTGACGCTCTGCT 1434
|||||
417 uPheIleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAlaA 434
|||||
1435 GTTTATCTGTGTGGAAGAAATGTGCTCACTCTGTACTCTTCTCATGTCAA 1484
|||||
434 rGAlaPheIleSerGlyGlyPheGlnAlaIaIaTyrValTyrThrProGlu 450
|||||
1485 GAGCGTTTATTTCTGGAGCTTTCAAGCGCATATGTTTACACACCTGAG 1534
|||||
451 ValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMe 467
|||||
1535 GTCTACCCACGCAACGGGGGCTTCGGCCTGGGCACCTGCAGCGCAT 1584
|||||
467 tAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuG 484
|||||
1585 GGCAAGAGTGGTGTCTCTCATCATCTCCGTTTCATCGCCAGGTGATGCTGG 1634
|||||
484 luSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeu 500
|||||
1635 AATCTCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTGCCCTCTG 1684
|||||
501 AlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyLe 517
|||||
1685 GCTGCCCTGCCCTCTCTTTTTCCTCATTTGAGACCAAGGCGGAGGACT 1734
|||||
517 uGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgGlyM 534
|||||
1735 GCAGGAGTCCAGCACCGGAGTGGGGCCAGGAGATGGTGGCGCGAGGAA 1784
|||||
534 ethisGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||
1785 TGCACGTGTCAGCTGTACCAGGTGCAACTCTGGCTCTCAGGAA 1828
|||||

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ49689

seq_documentation_block:

ID AAZ49689 standard; cDNA; 1906 BP.

XX

AC AAZ49689;

XX

DT 07-APR-2000 (first entry)
XX Rat organic cation transporter-like protein (OCT1p) partial cDNA.
DE
XX Human: organic cation transporter-like protein; OCT1p; transporter;
KW transmembrane; norepinephrine; neuroleptic; anticonvulsant;
KW antiparkinsonian; antidepressant; cellular process; cell proliferation;
KW screen; treatment; prevention; diagnosis; neurodegenerative disorder;
KW Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
KW CNS disorder; central nervous system; schizophrenia; depression;
KW behavioural; sleep disorder; eating disorder; rat; ss.
XX
OS Rattus sp.
XX
XX WO200000633-A1.
PN
XX 06-JAN-2000.
PD
XX 29-JUN-1999; 99WO-US14880.
PF
XX 30-JUN-1998; 98US-0107932.
PR
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX Goodearl AJ, Glucksmann MA;
PI WPI; 2000-137069/12.
DR
XX New nucleic acid encoding human organic cation transporter-like
PT protein, used for prevention, treatment and diagnosis of e.g.
PT neurological, behavioural or sleep disorders -
PT
XX Disclosure; Fig 4; 100pp; English.
PS
XX The present sequence is a partial cDNA encoding rat OCT1p (organic
CC cation transporter-like protein) and derived from rat frontal cortex
CC library. It was used to design primers to isolate OCT1p cDNA from human
CC foetal brain cDNA library. Human OCT1p is a member of the superfamily of
CC sugar and other transporter molecules that have 12 transmembrane
CC domains. It is highly expressed in brain tissue and has norepinephrine,
CC neuroprotective, neuroleptic, anticonvulsant, antiparkinsonian,
CC antidepressant activities. The OCT-like protein is used to
CC regulate a variety of cellular processes e.g. cell proliferation,
CC differentiation and survival. screen OCT1p modulators and detect mutation
CC in OCT1p gene. OCT1p modulators can be used to treat or prevent chronic
CC neurodegenerative disorders (e.g. Alzheimer's, Parkinson's,
CC Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic,
CC depression), behavioural, sleep and eating disorders.
XX
XX Sequence 1906 BP; 379 A; 547 C; 556 G; 424 T; 0 other;
SQ

alignment_scores:
Quality: 876.00 Length: 180
Ratio: 4.949 Gaps: 0
Percent Similarity: 98.333 Percent Identity: 93.889

alignment_block:
US-09-911-667A-2 x AA249689 ..

Align seg 1/1 to: AA249689 from: 1 to: 1906

369 SerGluGluAspTyrMetAspLeuLeuTrrpThrThrLeuSerGluPhePr 385
|||||
18 ACCAAGAGGATTACATGGACCTGCTGTGGACCCCTGCTGAGTTCCC 67

385 oGlyValLeuValThrLeuTrrpIleAspArgLeuGlyArgLysLysT 402
|||||
68 AGGTGCTCTGTGACTGCTGGGTCTATCGACCGCTGGCGCGCAAGA 117

402 hrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuPhe 418
|||||
118 CCATGCTCTGTGTTCTGCTATCTTTTCCCTGACGACCTCTCTGCTGTC 167

419 IleCysValGlyArgAsnValLeuThrLeuLeuLeuPheIleAlaArgAl 435
|||||
168 ATCTGCATTGGAAGAAATGTCTAACCCCTCTTACTGTTCATTGCAAGAGC 217

435 aPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluValT 452
|||||
218 GTTTATTTCGGAGGCTTCCAAGCAGCCTACGTTTACAGCCTGAGGTGT 267

452 YrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAla 468
|||||
268 ATCCAGGGCGACGAGGCGCTGGCCCTGGGCACCTGCAGCGCATGGCG 317

469 ArgValGlyAlaLeuLeuThrProPheIleAlaGlnValMetLeuGluSe 485
|||||
318 AGAGTGGCGCGCTCATCTACCTCCATTTCATAGCTCAGGTGATGCTGGAATC 367

485 rSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAlaA 502
|||||
368 TTCGCTGACTGACCTGGCCGCTCTACAGTGGCTGCTGCCCTCTCTGCTG 417

502 laLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGlyLeuGln 518
|||||
418 CCTTGGCTCTCTGCTTCTGCCCATCGAGACCAAGGCGGACGACTGCAG 467

519 GluSerSerHisArgGluTrrpGlyGlnGluMetValGlyArgGlyMetHi 535
|||||
468 GAGTCCAGGCCACCGGAGTGGGCCGAGGATGGTTGCCCGAGGACAAA 517

535 sGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||
518 CAGCACAGGCGTCCCGAGTGGAACTCTGGCTCTCAGGAG 557

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA45312

seq_documentation_block:
ID AAA45312 standard; cDNA; 401 BP.

AC AAA45312;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1887.

Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
expressed sequence tag; sEST; probe; chemotactic; proliferative;
immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
cerebroprotective; neuroprotective; antidepressant; gene therapy;
vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
central nervous system disorder; Alzheimer's disease; stroke;
Parkinson's disease; Huntington's disease; coagulation disorder;
haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

XX 15-OCT-1999; 99WO-US24206.

XX 15-OCT-1998; 98US-0104436.

XX (GEMY) GENETICS INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Bowman MR,

WPI: 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (sESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders -

Claim 1; Page 679; 803pp; English.

AAA45925 represent specifically claimed secreted expressed sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue sources. The sESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic; proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal; antiviral; antidiabetic; antiasthmatic; vulnerrary; antiulcer; osteopathic; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be used for gene therapy and in vaccines. The sESTs are useful as probes for the identification and isolation of full-length cDNAs and genomic DNA molecules which correspond to the sESTs. Proteins encoded by the sESTs are useful in assays for determining biological activity and raising antibodies. They may be useful for treatment of autoimmune disorders (multiple sclerosis, insulin dependent diabetes), allergic conditions (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers, osteoporosis, osteoarthritis, central nervous system disorders (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's disease), tumours, bacterial, fungal or viral infections, depression and psoriasis. AAA45926 to AAA45931 represent linker variants which are given in the exemplification of the present invention.

Sequence 401 BP: 69 A; 119 C; 104 G; 109 T; 0 other:

```

alignment_scores:
    Quality: 597.00      Length: 123
    Ratio: 4.975        Gaps: 0
    Percent Similarity: 97.561  Percent Identity: 96.748

alignment_block:
    US-09-911-667A-2 x AAA45312  ..

    Align seg 1/1 to: AAA45312 from: 1 to: 401

402 ThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPhe 418
|||||
33 ACCATGGCCCTGTGCTTTGTGCATCTTCTCCTCTCTGCAGCCTCCTGCTGTT 82
|||||
418 elleCysValGlyArgAsnValLeuThrLeuLeuLeuPheLeuAlaArgA 435
|||||
83 TATCTGTGTGGGAAGAAATGTCTCACTCTGTTTACTTCTCATATGC..AGAG 132
|||||
435 laPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluVal 451
|||||
133 CGTTTATTTCTGGAGGGCTTTTCAAGCGGCATATGTTTACACACCTCAGGTC 182
|||||
452 TyrProThrAlaThrArgAlaLeuLeuGlyLeuGlyThrCysSerGlyMetal 468
|||||
183 TACCCACAGGCACGCGGGGCCCTTGGGCTTGGGCACCTGCAGCGGGATGGC 232
|||||
468 aArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGluS 485
|||||
233 AAGAGTGGGTGCTCATCACTCCGTTTCATCGCCCAAGTGCATGCTGGAAAT 282
|||||
485 erSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAla 501
|||||
283 CCTCTGTGTACCTGCACCTCTGGCAGTTTACAGTGGGCTGCTGCCTCTCGGCT 332
|||||
502 AlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGlyLeuGlu 518
|||||
333 GCCCTGGCCCTCCTGCTTTTGGCCATTTGACACAAAGGCCGAGGACTGCA 382
|||||

```

518 nGluSerSerHisArgGlu 524
 |||||::: |||
 383 GGAGTCCAAGCCACTCGAG 401

seq_name: /SIDS2/qcdata/qeneseq/qeneseq/NA2001.DAT:AAH68534

seq_documentation_block:

ID AAH68534 standard; DNA; 309400 BP.

AA
AC
AAH68534:

26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEO ID NO: 7069.

AA Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
KW

OS *Corynebacterium glutamicum*.

PN EP1108790-A2

20-JUN-2001

18-DEC-2000 2000EP-0127688

XX
PR 16-DEC-1999. 99JP-0377484

PR 07-APR-2000; 2000JP-0139162.
PR 03-AUG-2000 2000JP-0280988

XX PA (KYOW) KY. A HAKKO KOGYO KK

XX
DT
Nakabayashi
Mitsunobu

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
yy

DR WPI; 2001-376931/40.

Novel polynucleotide

expression profile or pattern of a gene and identifying homologous gene

PS Disclosure: SEQ ID NO: 7069; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 309400 BP: 70133 A; 86477 C; 83115 G; 69675 T; 0 other;

alignment scores:

Quality:	498.50	Length:	480
----------	--------	---------	-----

Ratio:	1.701	Gaps:	14
--------	-------	-------	----

Similarity: 61.042 Percent Identity: 27.708

alignment_block:

US-09-911-667A-2 x AAH68534

Align seg 1/1 to: AAH68534 from: 1 to: 309400

90 ThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetIleLeuSerI 106

Tue Mar 12 15:54:26 2002

179148 TCAGGTATCGCGTGGCATTTGGATGGCATGGATGGATGGCGACTGATCTCTTTT 179197
106 eLeuAlaProClnLeuHisCysGluTyrTrpArgLeuProSerTrpGlnValA 123
179198 CGTCATGGCTGCGTTGGCCACTCAATGGGGCTATCCCTACTGNAACTT 179247
123 laLeuLeuThrSerValValPheValGlyMetMetSerSerSerThrLeu 139
179248 CCCTGCTCGGATCCATCGGTTTCGTGGATGGCCATCGCGGCTTCGCTG 179297
140 TrpGlyAsnIleSerAspGlnTyrClyArgAlaGlyThrGlyLeuLysIleSe 156
179298 GCGGGTTTGTGGCGGACAAGTTGGGCGTCCA.....CAGATTTT 179338
156 rValLeuTrpThrLeuTyrTyrGlyIle.....LeuSerAlaPheA 170
179339 TGCCTGTGCTTGTCTAGTTTATGGCTTGCCTGCCACTGGCGCGTCGGCGCTT 179388
170 laProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGly 186
179389 CAGTGTCTGCTGGCAATGTTGATGGCGTGCCTGCTGCTGGAGACTGCG 179438
187 IleGlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuPr 202
179439 CTGGCGCTGAACATCCCGCTTGCATCCACTCTGATTCCGAGTTTCTCC 179488
202 oMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpAlaI 219
236GlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLe 250
179589 GATAATGCTTGGCGTTGGCGTTAGCTCTTGGCTGCTGCTCCCTGCAATTA 179638
250 uPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValL 267
179639 CGCGGTGATGTCCGTCTCGCACTTCCAGAATCAGTACGTTTCTCTCGAGA 179688
267 euSerGlyAsnGlnGluLysAla.....IleAlaThrLeuLysArg 280
179689 AGAAAGCGCGCACCGACGAGCAAGCATGCTGTTGTTCTTCGAGAA 179738
281 IleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeuIleIleSe 297
179739 GCTCGCGTGGCGAAGGTAAAGCTGCCGATGCCACCGCTGGTGTCTCA 179788
297 rArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHisPheA 314
179789 TGACAACGCTGCCGAGGTTCCGTA...TCCATCTGGTCAGCTGCTTGC 179835
314 rgTrpTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSer 330
179836 GCAAGCGCACCGCTCGCGTGTGGATCGTGTGTTCTGTCATCAACTGTCTC 179885
331 TyrTrpGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVa 347
179886 TACTACGAGCGCTTCATTGGATTCACGCTGCTGGTTGCCAGCGGTTT 179935
347 lCysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuA 364
179936 CACCCTCGTGAAGTCT..... 179951
364 lAcCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThr 380
179952TCCAATTCACTCTGATCATCAC 179975
381 LeuSerGluPheProGlyValLeuValThrLeuTrpIleIleAspArgLe 397

```

179976 TTGGCTACGTTCCAGCGTATCGCGGTTCAGCGGTGTTGATTGAAGAAGTG 180025
      337 uGlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCysS 414
          |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
180026 GGGCGTGCAGCATATGGCC...ACGTTCTCGTGTGTTCTGCTATCT 180072
      414 erLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeuLeu 430
          || |||||::: |||| |::: |||| |::: |||| |:::
180073 CTGCAGCGCTCTAC.....GGCTGGCAAAATGTGAAGTGGCAGATC 180113
      431 PheIleAlaArgAlaPheIleSer.....GlyGlyPheClnA1 443
          |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180114 CTGTCGCGAGGTGTTACTATCTCTTCAACCTGGCGCATGGCGC 180163
      443 aAlaTy rValTy rThrProGluValTy rProThrAlaThrArgAlaLeuG 460
          || |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180164 ACTGTGCCAATATGGCGGGAGCTTTATCCCACATAATGTCCGTGGAAC 180213
      460 lyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrPro 476
          || |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180214 GAACGGTGTGCGCGGAGTTCGGCGCATGTCTCCATCATCGCTCG 180263
      477 PheIleAlaGInValMetLeuGluSerSerValTy rLeuThrLeu...Al 492
          |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180264 CTCATTGTTCCGCCAGTAGTCTTTTGGTGGACCAATGCTTTGTTTCG 180313
      492 avaly rSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuP 509
          |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180314 TCTCTTTGCCACCGGTTTCGATCGACGATTCGGCGGTTTACGCTGC 180363
      509 roileGluThrLysGlyGlyLeuGlnGlu..... 519
          || |||| |::: |||| |::: |||| |::: |||| |:::
180364 CT...GACCAGAGGTAACTCTCTCGTGATTAGTGAGATCCTTCCACC 180410
      520 .....SerSerHisArgGluTrpGlyGlnGluMetValGlyArgGly.M 534
          |||| |::: |||| |::: |||| |::: |||| |::: |||| |:::
180411 AGTTTTTCACCTGCCCAAGATGGTGCCTTTCTGCGCCTGGCGGAAGTCC 180460
      534 eHisGlyAlaGlyValThrArgSerAsnSerGlySer 546
          |||||::: |||| |::: |||| |::: |||| |::: |||| |:::
180461 TCCAATTCGGGGCAGGTCCGATACCGTGGGCACC 180498
seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH68260
seq_documentation_block:
seq_AA_H68260 standard; DNA; 1344 BP.
XX XX AC AAH68260;
XX XX
XX XX
26-SEP-2001 (first entry)
C glutamicum coding sequence fragment SEQ ID NO: 3295.
Corynebacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis; ds.
Corynebacterium glutamicum.
EPI108790-A2.
20-JUN-2001.
XX XX
18-DEC-2000; 2000EP-0127688.
XX XX
16-DEC-1999; 99JP-0377484.
XX XX
07-APR-2000; 2000JP-0159162.
XX XX
03-AUG-2000; 2000JP-0280988.
XX XX
(KYOW ) KYOWA HAKKO KOGYO KK.
XX XX
Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi
XX XX
Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX XX

```

seq name: /SIDS2/qcdata/geneseq/geneseq/NA2001.DAT:AAF67999


```

281 IleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeuIleIleSe 297
    |||:::  ::|||
764 GCTGCGGTGCGGAAGGTAAGCTGCGATGCCAGTCCACCGCTGGTTCA 813
    :::::  :::::
297 rArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHisPheA 314
    :::::  :::::
814 TGAACACGCTGCGGAGGGTCCGTA...TCCATCGGCAGCTGTTGCG 860
    |||:::  |||:::
314 rGrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAlaPheSer 330
    |||:::  |||:::
861 GCAAGCGCACCGTCGCGCTGTGGATCGTGTGTTCTGCATCAACTGTGCC 910
    |||:::  |||:::
331 TyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspVa 347
    |||:::  |||:::
911 TACTACGAGCGCTTCATTTGGATTCCATCGCTGGTGGCGACGGTTT 960
    :::::  :::::
347 lCysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuA 364
    :::::  |||
961 CACCCTCGGAAGTCT.....TTCGAATTCATCTCATCAACC 1000
    :::::  |||
364 lAcysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrThr 380
    :::::  |||
977 .....TTCGAATTCATCTCATCAACC 1000

381 LeuSerGluPheProGlyValLeuValThrLeuTrpIleIleAspArgLe 397
    |||:::  |||:::
1001 TTGGCTCAGCTTCCAGGCTATCGCGTTGACGCGTGTGATTGATAAAGTG 1050
    |||:::  |||:::
397 uGlyArgLysThrMetAlaLeuCysPheValIlePheSerPheCysS 414
    |||:::  |||:::
1051 GGGCGGTCGACACATTTGCC...ACGTTCTCTGGTGTCTCTCTATCT 1097
    |||:::  |||:::
414 erLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeuLeu 430
    |||:::  |||
1098 CTGACGCGCTCTAC.....GGCTTGGCAATGTGGAGTGGCGAGATC 1138
    |||:::  |||
431 PheIleAlaArgAlaPheIleSer.....GlyGlyPheGlnAl 443
    :::::  |||
1139 CTGGTCGAGGTGTTTACTATCTTCTTCAACCTGGCGCATGGGCGCG 1188
    |||:::  |||
443 aAlaTyrValTyrThrProGluValTyrProThrAlaThrArgAlaLeuG 460
    |||:::  |||:::
1189 ACTGTATGCCATTGGCGCGGAGCTTTATCCACTAATGTCCGTGGAAGTG 1238
    |||:::  |||:::
460 lLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrPro 476
    |||:::  |||:::
1239 GAACGGGTGTCGCGCGGATTCGGCGCATTGTCTCCATCATCGCTCCG 1288
    |||:::  |||
477 PheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeu...Al 492
    :::::  :::::
1289 CTCATTGTTCCGCCAGTGATTGCTTTTGGTGGACCAATGCTTTGTTCGC 1338
    |||:::  |||
492 aValTyrSerGlyCysCysLeuLeuAlaLeuAlaSerCysPheLeuP 509
    |||:::  |||:::
1339 TCTCTTCCACCGGTTTGGATTCGCGAGGATTCGGCGCATTCGCTTACGCTGC 1388
    |||:::  |||
509 roileGluThrLysGlyGlyLeuGlnGlu 519
    |||:::  |||
1389 CT...GACCAGAAGGGTAAGTCTCTCGCTGAT 1417

seq_name: /SID2/9cddata/geneseq/geneseq/NA2000.DAT-AAZ42410
seq_documentation_block:
ID AAZ42410 standard; cDNA; 480 BP.
XX
AC AAZ42410;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human 5' EST isolated from a cDNA library SEQ ID NO:169.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
gene therapy; chromosome mapping; upstream regulatory sequence;

```

```

forensic; location; development; protein synthesis; stability;
regulation; identification; ss.
Homo sapiens.
XX W09953051-A2.
XX 21-OCT-1999.
XX 09-APR-1999; 99WO-1B00712.
XX 09-APR-1998; 98US-0057719.
XX 28-APR-1998; 98US-0069047.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WP1; 2000-038446/03.
XX P-PSDB; AAY64796.
XX Novel secreted protein 5' expressed sequence tag sequences used in
XX diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX Claim 1; Page 265-266; 837pp; English
XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)
XX sequences, corresponding to human secreted proteins. AAY64651 to
XX AAZ43075 represent the EST-related proteins corresponding to AAZ42265 to
XX AAZ43075. The 5' ESTs can be used for producing secreted human gene
XX products. They can be used to identify and isolate 5' untranslated
XX regions (UTRs) and upstream regulatory regions which control the
XX location, development stage, rate, and quantity of protein synthesis, as
XX well as stability of mRNA. The ESTs are also useful as probes for
XX chromosome mapping, and to obtain full length cDNA clones. The ESTs can
XX also be used in forensic procedures to identify individuals, or in
XX diagnostic procedures to identify individuals having genetic diseases
XX resulting from abnormal gene expression. The products may also be used in
XX gene therapy protocols. The nucleic acids encoding signal peptides can be
XX used for directing extracellular secretion of a polypeptide or the
XX insertion of a polypeptide into a membrane, or importing a polypeptide
XX into a cell. The proteins encoded by the EST sequences may be useful in
XX treating a variety of human conditions. Secreted proteins have
XX therapeutic value, and the identification of new secreted proteins is
XX valuable. AAZ42249 to AAZ42264 and AAY64644 to AAY64650 represent
XX sequences used in the exemplification of the present invention.
XX Sequence 480 BP; 102 A; 115 C; 154 G; 100 T; 9 other;

alignment_scores:
Quality: 471.00 Length: 103
Ratio: 4.710 Gaps: 0
Percent Similarity: 97.087 Percent Identity: 95.146

alignment_block:
US-09-911-667A-2 x AAZ42410 ..
Align seg 1/1 to: AAZ42410 from: 1 to: 480
1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
|||||
166 ATGGAGGAGGACTTATTCCAGCTAAGGCAGTCCCGGTTGTGAATTCGG 215
|||||
17 gArgThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
|||||
216 TCGCACAGCGAGAGTGCAGAGTCCAGACGACGCGGCTTCAGGAGAGC 265
|||||
34 isGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
|||||
266 ATGAAGTCCAGATTGAAGGGTCCACGCTGGCGCTAGAGGCTGTGGAGCTG 315
|||||
51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
|||||

```

```

|||||.....|
316 GATGATGGGGCAKCTGTGCCCAAGAGTTGCCAATCCCAACCGATGATAC 365
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
|||||.....|
366 TTTCATGCTGGAAGATGCAAGTGGAGCCATTGGCTTTGGAAAAATTCAGT 415
84 rPlysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||.....|
416 GGAAGCTGTCTCTCTCAATGCTTGGCTTGATGCT.SATGCCATCGAG 464
101 MetMetIle 103
|||||
465 ATGATGATC 473

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ44679

seq_documentation_block:

ID_ AAZ44679 standard; cDNA to mRNA; 1901 BP.

XX AC AAZ44679;

DT 14-APR-2000 (first entry)

XX Rat liver anion transporter protein OAT2 encoding cDNA.

XX OAT2; anion transporter; liver; rat; dicarboxylic acid; prostaglandin;

XX non-steroid anti-inflammatory agent; anti-tumor; ss.

XX Rattus sp.

XX Key Location/Qualifiers

FT 51..1659

FT CDS /*tag= a

FT /product= "OAT2"

XX JP11346779-A.

XX 21-DEC-1999.

XX 03-JUN-1998; 98JP-0169174.

XX 03-JUN-1998; 98JP-0169174.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2000-1094/10.

XX P-PSDB; AAY51249.

XX A liver-specific org. anion transporter - can transport various

XX substances including dicarboxylic acids

XX Claim 4; Page 6-7; 13pp; Japanese.

XX This invention describes a novel rat liver-specific org. anion
 CC transporter OAT2. The liver-specific org. anion transporter OAT2
 CC can transport various substances including dicarboxylic acids,
 CC prostaglandins, non-steroid anti-inflammatory agents and anti-tumor
 CC agents. This sequence encodes the rat liver anion transporter OAT2
 CC protein described in the method of the invention.

XX Sequence 1901 BP; 346 A; 537 C; 571 G; 447 T; 0 other;

alignment_scores:

Quality: 446.50 Length: 552

Ratio: 1.493 Gaps: 18

Percent Similarity: 54.167 Percent Identity: 27.899

alignment_block:

US-09-911-667A-2 x AAZ44679 ..

Align seg 1/1 to: AAZ44679 from: 1 to: 1901

```

34 HisGluValGlnIleGlyValHisValGlyLeuGluAlaValGluLe 50
|||||.....|
189 CACATTGTGCCCTGCCGTGCCCTGCCAACCCTCAGTCACCGAGCTT 238
50 uAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspT 67
|||||.....|
239 ATGGCTGGAAGCCCATACCCCGGAG.....ACTGACGGCA 276
67 hrPheMet.....Val 70
|||||
277 GCTTTAGCTCTGCCTCGATTTGCCATCCCCAGACTGTCCCAATGTC 326
71 GluAspAlaValGluAlaIleGlyPheGlyLysPheGlnTrpLys....Le 86
|||||.....|
327 ACTTTGGGGACAGAGGTCTCCAACTCTGGGGAGGCTGAGGCTGAGCCCT 376
86 uSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetI 103
|||||.....|
377 CACGGTCCCTGCTCTCAGGGCTGGAGTACGACCGCTCAGAA..... 419
103 leLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSer 119
|||||.....|
420 .....TTCTCTCCACCATTTGCAACTGAGTGGGATCTTGTGTGT 458
120 TrpGlnValAlaLeu.....LeuThrSerVal....ValPheValGlyMe 133
|||||.....|
459 CAGCAGAGAGACTCAACAAATTTACGTCCACCTGCTTCTTCATTTGGTGT 508
133 tMetSerSerThrLeuTrpGlyValAsnIleSerAspGlnTrpGlyArgL 150
|||||.....|
509 GCTGGTGG:AGCCGTGGGTATGATACTTGTCTGACAGGTTTGCAGGC 558
150 ysThrGlyLeuLysIleSerValLeuTrpThrLeuTrpGlyIleLeu 166
|||||.....|
559 GCCGCTTCTGCTGCTGCTACGTGAGTCCCTGCTGCTGGGTGTGATG 608
167 SerAlaPheAlaProValTrpSerTrpIleLeuValLeuArgGlyLeuVa 183
|||||.....|
609 TCTGCAGGCTCCATCACTATCATCATGTCTGTAGTCACCCGTACACTCAC 658
183 lGlyPheGlyIleGlyValProGlnSerVal...ThrLeuTrpAlaG 199
|||||.....|
659 CGGCTCAGCCCTGGCTGCTTCAACCATCATTTGCTGCCACTGGAGTTGG 708
199 luPheLeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluVal 215
|||||.....|
709 AGTGGCTGGATGTGAGCACCCTGCTGGCCGGGTGTCATCAGCACCGTGC 758
216 PheTrpAlaIleGlyThrValPheGluValValLeuAlaValPheValMe 232
|||||.....|
759 TTCTGCTGS...GGAGCGGTGCTGCTGCTGGCAGCTGGTGGGTACCTGAT 805
232 tProSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuL 249
|||||.....|
806 CGGAGAGC.....TGGCGCTGGCTTCTGCTGGCTGCCACCTGCCGTGG 849
249 euLeuPheAlaValLeuLysPheTrpLeuProGluSerAlaArgTrpAsp 265
|||||.....|
850 TCCCAGGATCATCATCAGCATCTGGTGGTTCCTGAGTCTCAGCGTGGCTT 899
266 ValLeuSerGlyAsnGlnGlnLysAlaIleAlaThrLeuLysArgIleAl 282
|||||.....|
900 CTAACCCAGGGTGTGTGGAGGAGGCAAAAAAATACTTGTGAGCTGTGC 949
282 aThrGluAsnGlyAlaProMetProLeuGlyLysLeu..... 294
|||||.....|
950 CAAGCTCAATGGCGCGCGCTGGGTGGGTGAGGCGAGCTGAGCCAGGAGGCC 999
295 .....IleIleSerArgGlnGluAspArgGlyLysMetArg 306
|||||.....|
1000 TGAACAACGCTGCTCACCATGGAAGGGCGTTGCAAAAGACCCCTCATACTTA 1049

```

```

307 Aspleuphe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
||||| ||| : : : : : ||| : : : : :
1050 GACCTGTTCCGAACATCTCAGCTCGACATATCTCAGCTGCTCATGAT 1099
322 eiletrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuThr 339
: : : : : ||| : : : : : ||| : : : : :
1100 GGTGTGTTTGGAGTGAACCTTCTCTACTACGGCTGACTCTG..... 1142
339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
||||| |||
1143 .....GACGTGCTGGG..... 1154
356 AlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluLys 372
||||| : : : : :
1155 .....CTGGGCTGAACGT 1168
372 pTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuV 389
||||| ||| : : : : : ||| : : : : :
1169 GTACACAGACAGCTGCCTGTTTGGGGCTCTGGAGCTCCCTCCAAATTA 1218
389 alThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAla... 404
: : : : : ||| : : : : : ||| : : : : :
1219 TGGTCTACTTCTGTGGCGCTCTGGACGCCGTCTCAGGAGCTGG 1268
405 ...LeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheIleCy 420
||||| : : : : : ||| : : : : :
1269 ATGCTGTGGGCGCTCTGTACCTTTGGCACCGCTGCTGGTATCCTT 1318
420 sValGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheI 437
: : : : : ||| : : : : : ||| : : : : :
1319 GGAGCTAAGTATGATGATCAGTCTGTGTGGTGGGAAAGCTTTT 1368
437 leSerGlyGlyGlnAlaAlaTyrValTyrThrProGluValTyrPro 453
: : : : : ||| : : : : : ||| : : : : :
1369 CTGAAGCTCTTTTACTACGGCTTACCTGTTCACGTCCGAGTTGACCT 1418
454 ThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
||||| ||| : : : : : ||| : : : : :
1419 ACTGTGCTCAGACAGACAGAGTGGGACTTACTGCACTCATGGGAGGCT 1468
470 lGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGluSerSerV 487
||||| : : : : : ||| : : : : :
1469 AGGGGCTCTCTGGCCCCACTGGCGGC.....TTGCTGGATGGAG 1509
487 alTyrLeuThrLeu.....AlaValTyrSerGlyCysCysLeuLeuAla 501
: : : : : ||| : : : : : ||| : : : : :
1510 TGTGGCTGTGTGCTGCCAAAGTTGCTTACGGGGGATTGCCCTGTGGCT 1559
502 AlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyGluG1 518
||||| : : : : : ||| : : : : :
1560 GCTGCACCTGCACCTCTGCTGCT...GAGACGAAGAGGCACAGCTGCC 1606
518 nGluSer..... 520
1607 AGAGACCATCCAGAGTGTGGAGGAAGAGTACCCAGGAGGAGATGTGT 1656
521 .....SerHisArgGluTrpGly 526
||||| : : : : : ||| : : : : :
1657 AGGTCCGGGACTGAGTTGGACTAGGCACAGTTCTCCACAGGAGCTGGCA 1706
527 GlnGlu 528
1707 CAGAAG 1712

```

seq_name: /SIDS2/cgdata/geneseq/geneseq/NAL1996.DAT: AAT08702

seq_documentation_block:

ID AAT08702 standard; DNA; 1882 BP.

XX AAT08702;

XX 10-SEP-1996 (first entry)

XX

DE Rat OCT-1 gene.

XX KW Rat; OCT-1: transporter protein; cationic; xenobiotic; pharmaceutical;
 KW blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;
 KW proximal renal tubule cell; intestine; enterocyte; transgenic; renal;
 KW biliary; excretion; resorption; modulator; uptake; ss.

OS Rattus rattus.

XX Key Location/Qualifiers

FT CDS 38..1708

FT /*tag= a

FT /product= rat OCT-1

XX DE42424577-A1.

XX 18-JAN-1996.

XX 13-JUL-1994; 94DE-4424577.

XX 13-JUL-1994; 94DE-4424577.

XX (FARH) HOECHST AG.

XX Gorboulev V, Gruendeman D, Koepsell H;

XX WPI; 1996-069422/08.

XX P-PSDB; AAR77676.

XX Transporter protein for cationic xenobiotic(s) and pharmaceuticals,
 PT and related DNA and transformed cells - used e.g. to assess
 PT excretion and resorption of cationic cpds.

XX Claim 1: Fig 2A1; 13pp: German.

XX This is the nucleotide sequence of the rat OCT-1 gene encoding a
 CC new transporter protein able to transport cationic xenobiotics and
 CC pharmaceuticals from the blood into liver or kidney epithelial
 CC cells or from the intestine. The gene was isolated by injecting
 CC a rat kidney gene library into Xenopus laevis oocyte and isolating
 CC clones conferring uptake of 14C-tetraethylammonium. One clone
 CC designated OCT-1 was isolated. Expression of the gene was detect
 CC in proximal renal tubule cells, in liver epithelial cell and in
 CC intestinal enterocytes. The DNA can be used to generate transgenic
 CC cells for use in vitro test for renal/biliary excretion or
 CC intestinal resorption of xenobiotics and pharmaceuticals. The protein
 CC or cells expressing it can also be used to isolate modulators that
 CC block uptake of pharmaceutical by the renal tubules.

XX Sequence 1882 BP; 382 A; 510 C; 523 G; 467 T; 0 other;

alignment_scores:

Quality: 428.00 Length: 441

Ratio: 1.705 Gaps: 14

Percent Similarity: 56.916 Percent Identity: 29.478

alignment_block:

US-09-911-667A-2 x AAT08702 ..

Align seg 1/1 to: AAT08702 from: 1 to: 1882

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135

: : : : : ||| : : : : : ||| : : : : :

473 GCCTGGAAAGTGGACCTTTTCAGTCTCTGTGGAACCTTGGGCTTCTCCT 522

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152

: : : : : ||| : : : : : ||| : : : : :

523 GGGCTCCCTGTTGGGTTACATTGCAGACAGGTTGGCCGTAAGCTCT 572

152 lLeuLysIleSerValLeuThrLeuTyrTyrGlyIleLeuSerAla 168

||| : : : : : ||| : : : : : ||| : : : : :

573 GTCTCTTGTGTGACACGCTGTCACATCTGTGTCGGGTGTGCTAACAGCG 622

449	ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe	465
1400	GCTCAGCTGTACCCCTACATTCATCAGGAATCTTGGGATCATGGTATGCTC	1419
465	rGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValM	482
1450	TGCCCTGTGTGACCTGGGTGGATCTTCACCCCTTCATGGTGTTCAGGC	1499
482	eLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys	498
1500	TGATGGAAGTTGGCAAGCCCTGCCCTCATTTCTTTGGGGTTTTGGGC	1549
499	LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyG	515
1550	CTGACTGTCTGGGGCCATGACTCTTCTCTCCCA...GAGACCAAGGTGT	1596
515	yGlyLeuGlnGluSer.....Ser.HisArgGluTrpGly.GlnGluMe	529
1597	GGCTTTGGCTGAGACTATTGAAGAAGACAGAGAACCTGGGGAGAGAAAT	1646
529	tValGlyArgGlyMetHis	535
1647	CAAGGCCCAAGAAACAC	1665
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV49559		
seq_documentation_block:		
ID	AAV49559 standard; cDNA to mRNA; 1662 BP.	
XX		
AC	AAV49559;	
XX		
DT	21-OCT-1998 (first entry)	
XX		
DE	Human liver cell clone HP01293 cDNA #2.	
XX		
KW	Transmembrane domain; human; nutrition; cytokine; cell prol	
KW	differentiation; immune system; stimulator; suppressor; reg	
KW	hematopoiesis; activin; inhibitor; chemotactic; chemokinet	
KW	haemostatic; thrombolytic; ligand; anti-inflammatory; tumou	
XX		
OS	Homo sapiens.	
XX		
PN	W09821328-A2.	
XX		
PD	22-MAY-1998.	
XX		
PF	07-NOV-1997; 97MO-JP04056..	
XX		
PR	13-NOV-1996; 96JP-0301429.	
XX		
PA	(PROT-) PROTEGENE INC.	
PA	(SAGA) SAGAMI CHEM RES CENTRE.	
XX		
PI	Kato S, Kobayashi M, Sekine S, Yamaguchi T;	
XX		
DR	WPI; 1998-2979332/26.	
DR	P-PSDB; AAW64538.	
XX		
PT	Human protein having transmembrane domain - useful for, e.g. res	
PT	research and nutrition	
XX		
PS	Claim 3; Page 120-121; 205pp; English.	
XX		
CC	AAV49550-V49599 are cDNA sequences which encode human prote	
CC	a transmembrane domain. These proteins can be used for, e.g. 9-9	
CC	and nutrition, and may have cytokine and cell	
CC	proliferation/differentiation, immune stimulating/suppress	
CC	haematopoiesis regulating, tissue growth, activin/inhibin,	
CC	chemotactic/chemokinetic, haemostatic and thrombolytic, rec	
CC	anti-inflammatory or tumour inhibition activity.	
XX		
SQ	Sequence 1662 BP; 312 A; 486 C; 466 G; 398 T; 0 other;	


```
alignment_scores:
  Quality: 424.00      Length: 422
  Ratio: 1.767        Gaps: 11
  Percent Similarity: 56.872  Percent Identity: 29.858

alignment_block:
US-09-911-667A-2 x AAV49559

Align seg 1/1 to: AAV49559 from: 1 to: 1662

119 SerTrpGlnValAlaLeuThrSerValValPheValGlyMetMetse 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 TCCTGGAAGCTGACCTCTTTTCAGTCCCTGTTTGAATGCGGGCTCTCTT 482
135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
483 TGCTCTCTCGGTGGCTACTTTGCAGACAGGTTTGGCCGTAAGCTGT 532
152 lLeuLysIleSerValLeuTrpThrLeuTrpTyrglyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
533 GTCTCTGGGAAGCTGTGCTGCTCAACGGCGTGTGCGGCTGCTCATGGCC 582
169 PheAlaProValTyrsSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 TTCTCGCCCAACTACATGTCCATGTGCTTCCGCGCTGCTGCAGGGCT 632
185 e...GlyIleGlyValProGlnSerValThrLeuTyraAlaGluPheL 201
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
633 GGTGACGAAGGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 682
201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 TTGGCTCGGGCTCCAGGAAGCGGTGGCGATCATGCTACCAAGATTTG 732
218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProse 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733 ACGGTGGGCTGGTGGCGCTTACCGGGCTGGCC...TACGCCCTGCCTCA 779
234 rLeuGlyTrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 C.....TGGCGTGGCTGCAGCTGGCAGCTCCCTGCGCCACCTCTCTCT 823
251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyraPheValLeu 267
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
824 TCCTGCTCTACTACTGCTGTGCTGCGGAGTCCCTCGGTGGCTGTATCA 873
268 SerGlyAsnGlnLysAlaIleAlaThrLysArgIleAlaThrG 284
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
874 CAAAAAAGAAACACTGAAGCAATAAAGATAATGACACACATCGCTCAAAA 923
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
924 GAATGGGAAGTGGCTCTCTGCTGCTGCTTTAAAGATGCTTCCCTCGAAGAG 973
301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
974 ATGTCACCGAAAAAGCTGAGCCCTTCATTGTCAGACCTGTTCCGACGCG 1023
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1024 CGCCTGAGGAAGCGCACCTTCATCCTGTGATGCTGCTGCTGCTGCTGCTG 1073
328 aPheSerTyrglyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1074 TGTGCTCTATCAGGGGCTCATCTG.....CACATGG 1105
345 lYAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1106 GCGCCACACGCGGG..... 1119
362 SerLeuAlaCysGluTrpLeuSerGluGluAspTyraMetAspLeuLeuTr 378
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
1120 .....AACCTCTACCTGGATTTCCTTTA 1142
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleIleA 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 CTCCTGCTGCTGGAATCCCGGGCCCTTCATAGCCCTCATCACCATTG 1192
395 spArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePhe... 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1193 ACCGCGTGGCGCATCTACCCCATGGCGTGTCAAAATTTGTTGGCGGG 1242
411 SerPheCysSerLeuLeuLeuPheIle..... 419
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1243 GCAGCGTGCCTGCTCATGATGATTTTATCTCACCTGCACCTGCATGTA 1292
420 .....CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1293 CATCATATCATGCTGTGTTGGCCGAATGGGAATCACCATTGCAATA... 1338
432 leAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaValTyraThr 448
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1339 .....CAAATGATCTGCCTGTGTAAT 1359
449 ProGluValTyraProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1360 GCTGAGCTGTACCCACATTCGTGAGAACCTCGGAGTGTGTTGTTTC 1409
465 rGlyMetAlaArgValGlyAlaLeuIleThrPropheIleAlaGlnValM 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1410 CTCCTGTGTGACATAGGTGGGATAATCACCCCTTCATAGTCTTCAGGC 1459
482 etLeuGluSerSerValTyraLeuThrLeuAlaValTyraSerGlyCysCys 498
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1460 TGAGGAGGTCTGGCAAGCTTCCCTCATTTTGTTCGGGTGTTGGGC 1509
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyL 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1510 CTGCTGCGCGGAGTGACGTACTTCTTCCA...GAGACCAAGGGGT 1556
515 yGlyLeuGlnGluSer 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1557 CGCTTTGCCAGAGAC 1572

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:AAV49558

seq_documentation_block:
ID AAV49558 standard; cDNA to mRNA; 1888 BP.
XX
AC AAV49558;
XX
DT 21-OCT-1998 (first entry)
XX
DE Human liver cell clone HP01293 cDNA #1.
XX
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokinetic; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
OS Homo sapiens.
XX
FT Key Location/Qualifiers
CDS 90..1745
FT /*tag= a
FT /product= "transmembrane domain containing protein"
XX
PN W09821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX

```

Tue Mar 12 15:54:26 2002

PA (PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENTRE.
 XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;
 PI WPI; 1998-297932/26.
 DR P-PSDB; AAW64538.
 XX Human protein having transmembrane domain - useful for, e.g.
 PT research and nutrition
 XX Claim 4; Page 141-143; 205pp; English.
 XX AAV49550-V49599 are cDNA sequences which encode human proteins containing
 CC a transmembrane domain. These proteins can be used for, e.g. research
 CC and nutrition, and may have cytokine and cell
 CC proliferation/differentiation, immune stimulating/suppressing,
 CC haematopoiesis regulating, tissue growth, activin/inhibin,
 CC chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand,
 CC anti-inflammatory or tumour inhibition activity.
 XX
 SQ Sequence 1888 BP; 365 A; 545 C; 520 G; 458 T; 0 other;

alignment_scores:
 Quality: 424.00 Length: 422
 Ratio: 1.767 Gaps: 11
 Percent Similarity: 56.872 Percent Identity: 29.858

alignment_block:
 US-09-911-667A-2 x AAV49558 ..

Align seg 1/1 to: AAV49558 from: 1 to: 1888

```

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 TCCTGGAAGCTGGACCTCTTTCAGTCTCTTTGAATGGGGCTTCTTCT 571

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 TGGCTCTCTCGGTGTGGTACTTTGACAGAGGTTTGGCCGTAAAGCTGT 621

152 lLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
622 GTCTCTGGGAACGTGTGCTGGTCAACGGGCTCGGGGCTGCTCATGGCC 671

169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
672 TTCTCCGCCCACTACATGTCCATGCTGCTCTTCCGCTGTCTGCAGGGCT 721

185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
722 GGTCAACAAGGG/AACTGGATGGCTGGCTACACCTATACAGAAATTG 771

201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
772 TTGGCTCGGGCTCCAGAAAGAGGTTGGGATCATGTACAGATGGCCTTC 821

218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProSe 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
822 ACGGTGGGGCTGTGGCGCTTACCGGGCTGGCC...TAGGCCCTGCCTCA 868

234 rLeuGlyTyrArgTrpLeuIleLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
869 C.....TGGCGCTGGCTGCAGTGGCAGTCTCCCTGCCACCTTCTCT 912

251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
913 TCCTGTCTACTACTGTGTGTGGCGGAGTCCCTCGGTGGCTGTATCA 962

268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGl 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 CAAAAAAGAAACACTGAAGCAATAAAGATAATGGACCACATCGCTCAAA 1012

```

```

284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1013 GAATGGGAAGTTCCTCTCTCTGATTTAAAGATGCTTTCCTCGAAGAGG 1062

301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1063 ATGTCACCGAAAGAGTGCAGCCCTTCATTTGCAGACCTGTTCCGCACGCG 1112

312 HisPheArgTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1113 CGCTGAGGAAGCGCACCTTCATCTGATGTACCTGTGGTTCACGGACTC 1162

328 aPheSerTyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1163 TGTGCTCTATCAGGGGCTCATCTCTG.....CACATGG 1194

345 lYAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1195 GCGCACCGACGCGG..... 1208

362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1209 .....AACCTCTACCTGGATTTCCCTTTA 1231

378 pThrThrLeuSerGluPheProGlyValLeuValhrLeuTrpIleIleA 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1232 CTCGCTCTGTGCAAAATCCCGGGGCTTCATAGCCCTCATCACCATTG 1281

395 sPArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePhe... 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1282 ACCGCTGGCGCGCATCTACCCCATGGCCGTGCTCAAAATTTGTTGGCGGG 1331

411 SerPheCysSerLeuLeuLeuPheIle..... 419
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1332 GCAGCCTGCCTCGTCATGATTTTATCTACCTGCACCTGCACCTGGTTAA 1381

420 .....CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1382 CATCATATCATGTGTGTGGCCGAATCGGAATCACCATTGCAATA.... 1427

432 leAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThr 448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1428 .....CAAAATGATCTGCTGGTGAAT 1448

449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1449 GCTGAGCTGTACCCACATTCGTCAGGAACCTCGGAGTGTGCTGTTC 1498

465 rGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValM 482
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1499 CTCCTGTGTGACATAGGTGGGATAATCACCCCTTCATAGTCTTCAGGC 1548

482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1549 TGAGGGAGGTCTGGCAAGCCTTGCCCTTCATTTGTTGGCGTGTGGGC 1598

499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyL 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1599 CTGCTTGGCGGGAGTGACGCTACTTCTTCCA...GAGACCAAGGGGT 1645

515 yGlyLeuGlnGluSer 520
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1646 CGCTTTGGCAGAGACC 1661

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH57556
 seq_documentation_block:
 ID AAH57556 standard; cDNA; 4366 BP.
 XX
 AC
 XX
 DT 10-SEP-2001 (first entry)

XX Human brain cell specific cDNA sequence SEQ ID NO:396.
DE
XX Human: tissue specific; diagnosis; brain; heart; skeletal muscle;
KW lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX Homo sapiens.
OS
XX W0200132927-A2.
PN
XX 10-MAY-2001.
XX
XX 02-NOV-2000; 2000WO-US30396.
XX
XX 04-NOV-1999; 99US-0163508.
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Sornasse T, Sellhamer JJ, Watson GA;
PI
XX WPI; 2001-291057/30.
DR
XX
XX New cell and tissue specific polynucleotides useful for diagnosis,
PT prognosis or monitoring of treatments for disorders where the gene is
PT associated with a cancer, immunopathology or neuropathology -
XX
XX Claim 1; Page 309-310; 327pp; English.
XX
XX AAH57161 to AAH57576 represent cell and tissue specific polynucleotide
CC sequences (I). (I) can have cytostatic, immunomodulatory and
CC neuroprotective activities, and can be used in gene therapy. (I) and
CC assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC mimetics, peptides, proteins, agonists, antagonists, antibodies or
CC their fragments, immunoglobulins, inhibitors, drug compounds and
CC pharmaceutical agents. Expression of (I) in a sample indicates the
CC differentiation of embryonic stem cells into a tissue selected from
CC brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC tissues. (I) and (II) are used to produce an expression profile that
CC defines a metabolic or developmental process, treatment, condition,
CC disease or disorder. The gene profile can be used for diagnosis,
CC prognosis or monitoring of treatments and for investigating a
CC predisposition to a disorder where the gene is associated with a
CC cancer, immunopathology or neuropathology.
XX
XX Sequence 4366 BP; 916 A; 1187 C; 1195 G; 1068 T; 0 other;

alignment_scores:
Quality: 413.00 Length: 710
Ratio: 1.237 Gaps: 22
Percent Similarity: 47.042 Percent Identity: 22.254

alignment_block:

US-09-911-667A-2 x AAH57556 ..

Align seg 1/1 to: AAH57556 from: 1 to: 4366

21 GluSerAlaArgSerGluAspThrAlaSerGlyGluHisGluValG1 37
660 GAAGGTGGTCCAGTATCTACTAGGCGCCATGACGAGGATGATGA 709
37 nile...GluGlyValHisValGlyLeuGluAlaValGlu..... 49
710 GATCTATGAAGGGAATATCAGGCGATTCCCGGGCAGAGTCTGGGGCA 759
49
760 AAGCGGACGGATGCGATGGGCGCCCTCGGTGGAGTAGGGGGGCG 809
50 LeuAspGlyAlaAlaValPro..... 57

810 TTGAGTGTGGGGAGGGTCCCTCTGGGGCGGGGGAGGCACACGACG 859
58LysGluPheAlaAsnProThrAspThrPheMetValG 71
860 GAAAGAACGAGAAGAACTGGCCCAA.....CAGTATG 891
71 LuAspAlaValGluAlaTleGlyPheGlyLysPheGlnTrpLysLeuSer 87
892 AAGCCATCCTACGGGAGTGTGGCCACGCGCTTCAGTGGACACTGTAT 941
88 ValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMetIleLe 104
942 TTGTGCTTGTCTGGCGCTGATGGTGTGACGGTGTGGAGGTCTTTGTGGT 991
104 uSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpG 121
992 GGGCTTCGTGCTGCCAGCGCTGAGAAACACATGTGCTGTCGACGTCCA 1041
121 InValAlaLeuLeuThrSerValValPheValGlyMetMetSerSer 137
1042 ACAAGGCGATGTAGCCCTCATCTGTACCTGGGCATGATGGTGGAGCC 1091
138 ThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGlyLeu 154
1092 TTCTCTGGGGAGGTCTGGCTGACCGCTGGGTGGGAGCGAGTGTCTGCT 1141
154 sIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAlaPheAlaP 171
1142 CATCTCGCTCTCAGTCAACAGCGCTCTGCTCTCTCTCTCTCTCTCT 1191
171 roValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPheGlyIle 187
1192 AGGTTTACGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1241
188 GlyGly...ValProGlnSerValThrLeuTyrAlaGluPheLeuPro 203
1242 GGAGGTGTCATCCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1291
203 tIysAlaArgAlaLysCysIleLeuLeuGluValPheTrpAlaIleG 220
1292 GGAGAAACGAGGGGACATTTGAGTGGCTCTGCTGCTGTTTGGATGAT 1341
220 LyThrValPheGluValValLeuAlaValPheValMetProSerLeuGly 236
1342 GTGCGGTGTACGAGCTGTCTATGGCGCTGGGCCCATCATCCCCACTAT 1391
237TrpArgTrpLeu 240
1392 TGGAGTTTTTCAGATGGGTCTTCCCTACCAGTTCCACAGCTGGAGGTCT 1441
240 uLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCysPhe 257
1442 CGTCTCTGCTCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1491
257 rpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGluLys 273
1492 CGCAGCGCTGAGAGCCCGCTTCTCTCTAGAGAATGAAAGCATGATGAG 1541
274 AlaIleAlaThrLeuLysArgIleAlaThrGluAsn.....G1 286
1542 GCCTGGATGGTCTGAAGCAGGTTCATATACCAACATCGGAGCCAAAGG 1591
286 yAlaPro.....MetProLeuGlyLysLeuIleLeuSerArgGlnG 300
1592 ACATCTCTGAGCGAGTGTCTCTCAGTAACCCACATTAAGACGATTCAT 1641
300 LuAspArg..... 302
1642 AGGATGAATTGATGATGATCCAGTCGACACACAGGGACCTGGTACCAGCG 1691
303GlyLysMetAr 306

```

498 sLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyG 515
      :|||      |||||      |||||      |||||      |||||
2633 CCTGAAG.....CTGCCT...GAGACCGGGGGC 2658
      |||||
515 lyGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGly 531
      |||||
2659 AGGTGCTGCAG.....TGA 2672

532 ArgGlyMetHisGlyAlaGlyValThrArg 541
      |||||      |||      |||||      |||
2673 AGGGTCTCTAGGCGTTTGGGATTGGCAGG 2702

seq_name: /sids2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC85822
seq_documentation_block:
ID AAC85822 standard; DNA; 2121 BP.
XX
AC AAC85822;
XX
06-AUG-2001 (first entry)
DT
XX
hoAT3 DNA.
DE
XX
Human; organic anion transporter; hoAT; liver; kidney;
KW membrane protein; transport; organic anion; splice variant;
KW
XX
Homo sapiens.
OS
XX
Key Location/Qualifiers
FH 62..1690
FT CDS
FT /*tag= a
FT /product= "hoAT3"
XX
WO200104283-A2.
XX
18-JAN-2001.
PD
XX
12-JUL-2000; 2000WO-US18980.
XX
12-JUL-1999; 99US-0143771.
XX
(META-) METABASIS THERAPEUTICS INC.
XX
Sun W;
PI
XX
WPI; 2001-367057/38.
DR
P-PSDB: AAB47274.
XX
Nucleic acids encoding human organic anion transporter poly-
useful in gene therapy procedures -
XX
Claim 1; Fig 4; 95pp; English.
XX
The sequences given in AAC85819-24 encode human organic anion
transporter (hoAT) polypeptides. hoAT polypeptides are pro-
expressed in the liver and kidneys of humans. hoAT's are me-
proteins that facilitate the transport of organic anions acro-
cell membrane. The mechanism of transport is thought to be
or tertiary active transport involving exchange of another
hoAT2A and hoAT2B are thought to be splice variants as they
identical except at the C-terminal end. hoAT proteins and
encoding them, may be used in the prevention, treatment and
of diseases associated with inappropriate hoAT expression.
XX
Sequence 2121 BP; 445 A; 680 C; 522 G; 474 T; 0 other;
SQ

alignment_scores:
Quality: 407.00 Length: 556
Ratio: 1.443 Gaps: 17
Percent Similarity: 50.719 Percent Identity: 23.741

```

alignment_scores:		
Quality:	407.00	Length: 556
Ratio:	1.443	Gaps: 17
Percent Similarity:	50.719	Percent Identity: 23.741

alignment_block:

US-09-911-667A-2 x AAC85822

Align seg 1/1 to: AAC85822 from: 1 to: 2121

```
62 AsnProThrAspThrPheMetValGluAspAlaValGluAlaIleG1 78
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
44 GACCGGCCACCACTGCCATGACCTTCTCGGAGATCCTGGACCGTGTGG 93
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
78 y...PheGlyLysPheGlnTrpLysLeuSerValLeuThrGlyLeuAla 94
   |  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
94 AAGCATGGGCCATTTCCAGTCTCCATGCTAGCCATAGCTAGGCGCTCCCGA 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
94 rp.....MetAlaAspAlaMetGluMetMetIleLeuSerIleLeuAla 108
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
144 TCCTCAACATGGCCCAACCAACCTGCTGCAGATCTTCACAGCGCGCAC 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
109 ProGlnLeuHisCysGlu.....TrpAr 116
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
194 CTGTCCAGCACTGTGCGCGCGGCCCAACATGCTTCACAGGCGCTTGGGT 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
116 gLeuPro..... 118
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
244 GCTCCCATGGGCCCAAAATGGGAAGCTGAGAGGTGCTCCGTTTGTAC 293
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
118 ..... 118
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
294 ATCCGCCCAATGCCAGCTTGCCCAATGACACCCAGAGGCCATGGAGCCA 343
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
119 .....Se 119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
344 TGCTGGATGGGTGGTCTACACACGACCAAGGACTCCATTTGTGACAGA 393
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
119 rTrpGlnValAla.....LeuLeuThrServ 128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
394 GTGGACTTGGTGCACCTCAACAACTGAAGGAGATGGCCCACTCA 443
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
128 alValPheValGlyMetMetSerSerThrLeuTrpGlyAsnIleSer 144
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
444 TCTTCATGGCAGGTACTGATTTGGAGGCTCGTGTGGAGACCTGTCT 493
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
145 AspGlnTrpGlyArgLysThrGlyLeuLysIleSerValLeuTrpThrLe 161
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
494 GACAGTTTGGCGCGAGGCCCTCTGACCTGCAGCTACCTGCTGTGCGC 543
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
161 uTyTrpGlyIleLeuSerAlaPheAlaProValTyTrpTrpIleLeuV 178
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
544 AGCCAGCGGCTCGGTGCAGCCTTCAGCCCACTTCCCATCTACATGG 593
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
178 alLeuArgGlyLeuValGlyPheGlyIleGlyValProGlnSer... 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
594 TCTTCCGCTTCTGTGTGGCTTTGGCATCTCAGGCATATACCCTGAGCAC 643
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
194 ValThrLeuTyAlaGluPheLeuProMetLysAlaArgAlaLysCysI1 210
   ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
644 GTCATCTGAATGTGGAATGGGTGGCTACCCGGATGCGGGCCATCATGTC 693
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
210 eLeuLeuIleGluValPheTrpAlaIleGlyThrValPheGluValVall 227
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
694 GACAGCACTCGGTACTGCTACACCTTTGGCCAG..... 727
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
227 euAlaValPheValMetProSerLeuGly.....TrpArg 238
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
728 .....TTCATTCTGCCCGGCTGGCTACGCCATACGCCATCCCGCAGTGGCGT 769
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
239 TrpLeuLeuIleLeuSerAlaValProLeuLeuPheAlaValLeuCy 255
   ||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
770 TGGCTGCAGTTAACTGTGTCCATCTCTCTCTCTCTCTCTCTCTCTCTCT 819
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 sPheTrpLeuProGluSerAlaArgTyTrpAspValLeuSerGlyAsnGln 272
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
820 CTGGTGACACACAGATGCCATAGCTGGTGGTCTTGTCTGTGGAAGTCCT 869
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAAL6697

seq_documentation_block:

ID AAAL6697 standard; cDNA; 3580 BP.

XX

157 alLeuTrpThrLeuTyTyTcGlyileLeuSerAlaPheAlaProValTyTyr 173
|||
589 ATCTCTTGTGCGCAGCGAGTGGCTCCAGTCTGCTGCTTCAGCCCCAGCGCTC 638
:
174 SerTrpIleLeuValLeuArgGlyLeuValcIlyPheGlyIleGlyGlyVa 190
:
639 ACTGCTATATGATCTTCCGATTCTCTGTGGCTGCAGCATCTCGGGCAT 688
:
190 lProGluSer...ValThrLeuTyTyAlaGluPheLeuProMetLysAlaA 206
:
689 TTCCTCTGAGCACCATTATCTTGAATGTGGAAATGGGTACCCACCATCCACGC 738
:
206 rgAlaLysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrVal 222
:
739 GGGCCATCTCATCAACAACATTATTGGGTACTCTTACACCATTTGGTCAA... 785
:
223 PheGluValValLeuAlaValPheValMetProSerLeuGly... 236
:
786TTCATTCTGCTGGCTGGCTTACGCTTCTTCAATCT 814
:
237TrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuP 251
:
815 TCCWTAGTGGCGCTGGCTACAGTTGTCGCTGCTGCTTCTTCAATCT 864
:
251 heAlaValLeuCysPheTrpLeuProGluSerAlaIleArgTyAspValLeu 267
:
865 TCTCTTGTCTGCTGCTGGTACCAGAGTCCATACGCTGGCTGGTCTGTCG 914
:
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThr... 283
:
915 TCTGAAATTTCTCACGAGCTCTGAAGCACTCAACAGTTGGGTACGCTT 964
:
284GluAsnGlyAlaProMetProLeuGlyLysLeuIleI 296
:
965 CAACGGCAAGAGCAGGAAGGGGAAAAGCTCACTGTGGAGGAGCTGAAGT 1014
:
296 leSerArgGlnGluAsp.....ArgGlyLysMetArg 306
:
1015 TCAACTTGCAGAGGACATCACCTCAGCCAAAGGTCAATATGGCTTATCT 1064
:
307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
:
1085 GACTTGTCCAGAGTCCATCTCTCGCGGTGACCTTCTCTCTCTCTCT 1114
:
322 eileTrpPheSerAsnAlaPheSerTyTyTyGlyLeuValLeuLeuThrT 339
:
1115 GGCCCTGGTTTCTACTGGCTTGGCTACTACAGTTTGGCTTACAGTGGAGTAG 1164
:
339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
:
1165 AAGAAATT.....GGAGTCAACATC..... 1184
:
356 AlaValGluAlaLysCysSerLeuAlaCysGlnTyTrLeuSerGluGluAs 372
:
1185TACATA..... 1190
:
372 pTyMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuV 389
:
1191CTCCAGATCATCTCGTGGGGT...GACATTCCGCCCAAGTTCA 1233
:
389 alThrLeuTrpIleAlaAspArgLeuGlyArgLysLysThrMetAlaLeu 405
:
1234 TCACAATCTCTCCATAAGTTATCTGGCGCGGCCATCATCAGGGGCTC 1283
:
406 CysPheValIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
:
1284 CTCCTGATCTCGCAGGAGTGGCCATCTCGGCCCTCATCTTTGTGTCTTC 1333
:
421 lGlyArgAsnValLeu...ThrLeuLeuLeuPheIleAlaArgAlaPheI 437
:
1334 AGAAATG...AGCTCTTTGAGAAGCAGCAGCTGGCTGTATTTGGGAAGGATGCC 1383
:
437 leSerGlyGlyPheGlnAlaAlaTyValTyThrProGluValTyPro 453

157 alLeuTrpThrLeuTyTyTcGlyileLeuSerAlaPheAlaProValTyTyr 173
|||
589 ATCTCTTGTGCGCAGCGAGTGGCTCCAGTCTGCTGCTTCAGCCCCAGCGCTC 638
:
174 SerTrpIleLeuValLeuArgGlyLeuValcIlyPheGlyIleGlyGlyVa 190
:
639 ACTGCTATATGATCTTCCGATTCTCTGTGGCTGCAGCATCTCGGGCAT 688
:
190 lProGluSer...ValThrLeuTyTyAlaGluPheLeuProMetLysAlaA 206
:
689 TTCCTCTGAGCACCATTATCTTGAATGTGGAAATGGGTACCCACCATCCACGC 738
:
206 rgAlaLysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrVal 222
:
739 GGGCCATCTCATCAACAACATTATTGGGTACTCTTACACCATTTGGTCAA... 785
:
223 PheGluValValLeuAlaValPheValMetProSerLeuGly... 236
:
786TTCATTCTGCTGGCTGGCTTACGCTTCTTCAATCT 814
:
237TrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuP 251
:
815 TCCWTAGTGGCGCTGGCTACAGTTGTCGCTGCTGCTTCTTCAATCT 864
:
251 heAlaValLeuCysPheTrpLeuProGluSerAlaIleArgTyAspValLeu 267
:
865 TCTCTTGTCTGCTGCTGGTACCAGAGTCCATACGCTGGCTGGTCTGTCG 914
:
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThr... 283
:
915 TCTGAAATTTCTCACGAGCTCTGAAGACCTCAACAGTTGGGTACGCTT 964
:
284GluAsnGlyAlaProMetProLeuGlyLysLeuIleI 296
:
965 CAACGGCAAGAGCAGGAAGGGGAAAAGCTCACTGTGGAGGAGCTGAAGT 1014
:
296 leSerArgGlnGluAsp.....ArgGlyLysMetArg 306
:
1015 TCAACTTGCAGAGGACATCACCTCAGCCAAAGGTCAATATGGCTTATCT 1064
:
307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
:
1085 GACTTGTCCAGAGTCCATCTCTCGCGGTGACCTTCTCTCTCTCTCT 1114
:
322 eileTrpPheSerAsnAlaPheSerTyTyTyGlyLeuValLeuLeuThrT 339
:
1115 GGCCCTGGTTTCTACTGGCTTGGCTACTACAGTTTGGCTTACAGTGGAGTAG 1164
:
339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
:
1165 AAGAAATTT.....GGAGTCAACATC..... 1184
:
356 AlaValGluAlaLysCysSerLeuAlaCysGlnTyTrLeuSerGluGluAs 372
:
1185TACATA..... 1190
:
372 pTyMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuV 389
:
1191CTCCAGATCATCTCGTGGGGTT...GACATTCCGCCCAAGTTCA 1233
:
389 alThrLeuTrpIleAlaAspArgLeuGlyArgLysLysThrMetAlaLeu 405
:
1234 TCACAATCTCTCCATAAGTTATCTGGCGCGGCCATCATCAGGGGCTC 1283
:
406 CysPheValIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
:
1284 CTCCTGATCTCTGGCAGGAGTGGCCATCTCGGCCCTCATCTTTGTGTCTTC 1333
:
421 lGlyArgAsnValLeu...ThrLeuLeuLeuPheIleAlaArgAlaPheI 437
:
1334 AGAAATGTCCTTTGAGAACAGCAGCTGGCTGTATTTGGGAAGGATGCC 1383
:
437 leSerGlyGlyPheGlnAlaAlaTyTyValTyThrProGluValTyPro 453

Tue Mar 12 15:54:26 2002

```
:::|||||:::|||||:::  :::|||||  |||:::|||||
1384 TGTCTGGCTCCTTCAGCTGCCCTCTTCCTCTACACAAGTGAGCTCTACCCCT 1433
454 ThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
|||:::  |||  |||:::|||||  :::  |||
1434 ACAGTCCTCAGCAACAGGTATGGGTATCAGTAACATATGGGCTCGAGT 1483
470 lGlyAlaLeuIleThrProPheIle.....AlaGlnValM 482
|||:::|||||:::|||||:::  |||  :
1484 GGAAGTATGATAGCCCACTGCTGAAAATCAGGGGAGAACTGCAGCCCT 1533
482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
:::  :::  :::  |||:::|||||:::  :::  |||
1534 TCATCCCTAATGTCATCTTTGGACCATGACTCTACTGGGAGGC..... 1577
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyG 515
:::  :::  :::  |||:::|||||  :::  :::
1578 .....AGTGTGCCTTCTTTCTGCTTGAGACCCCTCAATCG 1612
515 yGlyLeuGlnGluSerSerHis.....ArgGluTrpGlyGlnGlu 528
|||  |||:::  :::  :::  |||  |||:::
1613 GCCCTTACCAGAAAATATCGAGGACATACAAGACTGGTACCACCAA 1658
```

/cgn2_6/prodata/2/ina/5A_COMB.seq	:US-08-337-339-7 +	116.50	189.14	0.0069	6513
/cgn2_6/prodata/2/ina/5A_COMB.seq	:US-08-724-095-7 +	116.50	189.14	0.0069	6513
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq	:PCT-US95-14262-7 +	116.50	189.14	0.0069	6513
/cgn2_6/prodata/2/ina/PCTUS.COMB.seq	:PCT-US95-14378-7 +	116.50	189.14	0.0069	6513
/cgn2_6/prodata/2/ina/6A_COMB.seq	:US-09-024-020B-7 +	116.50	189.06	0.0070	6556

Command line parameters:

```

MODEL=frame_p2n.model -DEV xlp
-Q=/cgn2_1/USPro_spool/US09911667/runat_07032002_130318_29553/app_query.fastq_1.615
-DB=IssuedPatents_NA -QFWT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsm62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pts
-NORM-ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09911667 @CGN1_186 -NCP=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

```

Search information block:

```

Search information request:
Query: US-09-911-667A-2
Query length: 548
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 11238999
Search time (sec): 89.480000

```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-501-572-4 +	430.00	884.62	1.3e-41	1882 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-040-444-4 +	430.00	884.62	1.3e-41	1882 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-501-572-6 +	426.00	875.86	3.9e-41	1896 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-040-444-6 +	426.00	875.86	3.9e-41	1896 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-501-572-5 +	425.50	874.85	4.4e-41	1885 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-040-444-5 +	425.50	874.85	4.4e-41	1885 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-647-397-1 +	401.00	820.34	4.8e-38	2102 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-09-033-840A-1 +	230.00	347.79	1.0e-11	4411529 ?
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-964-127-1 +	203.50	390.56	4.2e-14	2460 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-09-103-840A-2 -	166.00	209.22	0.0005	4403765 ?
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-09-103-840A-1 -	158.00	191.87	0.0049	4411529 ?
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-09-591-025-8 +	154.50	283.75	3.7e-08	2592 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-125-468-1 -	151.50	244.52	5.7e-06	30001 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-474-933-3 +	151.50	244.52	5.7e-06	30001 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-928-692-9 +	147.00	265.55	3.8e-07	3000 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-09-031-392-1 +	145.00	264.53	4.4e-07	2343 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-299-549-1 +	145.00	264.53	4.4e-07	2343 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-994-035C-4 -	142.00	243.03	6.9e-06	7198 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-494-907-13 -	138.00	257.57	1.1e-06	1269 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: PCT-US96-10986-13 -	138.00	257.57	1.1e-06	1269 ?
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-494-907-2 +	138.00	239.03	1.2e-05	5076 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: PCT-US96-10986-2 -	138.00	239.03	1.2e-05	5076 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-494-907-4 +	138.00	236.42	1.6e-05	6170 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: PCT-US96-10986-4 -	138.00	236.42	1.6e-05	6170 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-494-907-83 -	138.00	235.96	1.7e-05	6387 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: PCT-US96-10986-3 -	138.00	235.96	1.7e-05	6387 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: PCT-US96-12545-5 +	134.00	229.57	3.9e-05	5390 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-964-127-5 +	133.00	245.32	5.1e-06	1411 !
/cgn2_6/ptodata/2/ina/5B COMB. seq: US-08-882-704A-4 +	131.50	219.31	0.0001	7742 !
/cgn2_6/ptodata/2/ina/5A COMB. seq: US-07-925-695-3 +	129.00	222.83	9.2e-05	3970 !
/cgn2_6/ptodata/2/ina/5A COMB. seq: US-07-925-695-2 +	129.00	211.04	0.0004	9589 !
/cgn2_6/ptodata/2/ina/5A COMB. seq: US-07-925-695-2 +	129.00	211.04	0.0004	9589 !
/cgn2_6/ptodata/2/ina/6B COMB. seq: US-08-964-127-3 +	127.00	231.60	3.0e-05	1490 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-931-999-4 -	125.50	208.14	0.0006	6755 !
/cgn2_6/ptodata/2/ina/5B COMB. seq: US-08-850-880-11 +	119.50	181.20	0.0192	19182 !
/cgn2_6/ptodata/2/ina/5B COMB. seq: US-08-944-916-11 +	119.50	181.20	0.0192	19182 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: PCT-US96-05320A-899 +	118.00	211.76	0.0004	15182 ?
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-024-020B-1 +	116.50	190.29	0.0060	5977 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-09-024-020B-2 +	116.50	190.22	0.0060	6007 !
/cgn2_6/ptodata/2/ina/6A COMB. seq: US-08-338-702-7 +	116.50	189.14	0.0069	6513 !

alignment_scores:		
Quality:	430.00	Length: 441
Ratio:	1.713	Gaps: 14
Percent Similarity:	56.916	Percent Identity: 29.478

alignment block:

```

US-09-911-667A-2 x US-08-501-572-4
Align seg 1/1 to: US-08-501-572-4 from: 1 to: 1882

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSer 135
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
473 GCCTGGAAAGTGGACCTTTTTCAGTCTCTGTGGAACCTTGGGCTTCTTCT 522
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
523 GGCTCCCTGCTGGTGGTTCATTGCACAGAGTTTGGCCGTAACTCT 572
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::

152 lyLeuTyIleSerValLeuThrTrpLeuTyrGlyIleLeuSerAla 168
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::
573 GTCTCTGGTGCACCACTGCTGCATCTGTCTCGGTGTGCTTAACAGG 622
   :::::|||||:::|||||:::|||||:::|||||:::|||||:::

```

449 ProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSe 465
1400 GCTGAGCTGTACCTACATTCATCAGGAATCTTGGATGATGATGCTC 1449
465 rGlyMetAlaAqValGlyAlaLeuLeuThrProPheIleAlaGlnValM 482
1450 TGCCCTGTGTGACCTGGTGGGATCTTACCCCTTCATGCTGTTGAGGC 1499
482 etLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCys 498
1500 TGATGAAGTTTGGCAAGCCTGCCCTCATTTTGTGGGTTTGGGC 1549
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGI 515
1550 CTGACTGCTGGGCCATGACTCTTCTTCCCA...GAGACCAAGGTGT 1596
515 yGlyLeuGlnGluSer.....Ser.HisArgGluTrpGly.GlnGluMe 529
1597 GGCTTTGCTGACACTATTGAACAAGCAGAACTGGGAGGAGGAAAT 1646
529 tValGlyArgGlyMethis 535
1647 CAAAGGCCAAAGAAACAC 1665

seq_name: /cgn_2_6/ptodata/2/ina/6A_COMB.seq:us-09-040-444-4

seq_documentation_block:
; Sequence 4, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-040-444-4

alignment_scores:
Quality: 430.00 Length: 441
Ratio: 1.713 Gaps: 14

169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
623 GTGCCCCAGACTATACATCATCTTTCGCTGCTGCGGCGAT 672
185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaClyPheL 201
673 GGTACGACGAGGCGAGCTGGTGGCTATACCTTGATCACAGAGTTTG 722
201 euProMetLysAlaAraAlaLysCysIleLeuLeuIleGluValPheTrp 217
723 TCGGCTCTGGCTACAGAGAACACGCGCCATTTGTACCATGAGCTTC 772
218 AlaIleGlyThrValPheGluValAlaValPheValMetProSe 234
773 ACAGTGGGCTAGTGGGCTTGGCGGGTGGCC...TATGCCATTCAGA 819
234 rLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeup 251
820 C.....TGGCGCTGGCTCAGCTAGCTGTCTCCCTGCTACCTCTCT 863
251 heAlaValLeuLysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
864 TCCTGCTGATTACTGGTTGTCCAGAAATCCCGGCTGGCTGTCTGTC 913
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGI 284
914 CAGAGACACACCGCGAGCTGTCCAGGATAATGAGCAAAATGACACAGA 963
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
964 GAACGGGAAGTGGCTCTGCTGACCTGAAGATGCTCTGCTTGGAGG 1013
301 sPArgGlyLysMetArg.....AspLeuPhe...ThrPro 311
1014 ATGCCCTCAGAAAGGAGTCTTTCGTTGGCGACCTTCCGCACTCC 1063
312 HisPheArgTrpThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
1064 AACCTGAGGAGACACACCGCTATCTATGTTCTCTGTGTC 1113
328 aPheSerTyrTyrGlyLeuValLeuThrThrGluLeuPheGlnAlaG 345
1114 TGTGCTGTACAGGCTCTCATCAG..... 1138
345 lyAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
1139CACGTGGGAGCCACAGGGGCCAACCTC 1165
362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
1166TACCTGGACTCTTTTAA 1182
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleLeA 395
1183 TTCTCTCTGGTGAATCCCGCGGCTTCATCATCTGCTGCTACCATG 1232
395 sPArgLeuGlyArgLysThrMetAlaLeuCys....PheValIlePhe 410
1233 ACCGATTTGGCGCATCTACCAATAGCGGCTCGAATCTGGTACGGGG 1282
411 SerPheCysSerLeuLeuLeuPheIle..... 419
1283 GCAGCTGCTCTCTCATGATCTTTATCCCGCATGAGCTGCATGCTGAA 1332
420CysValGlyArgAsnValLeuThrLeuLeuLeuPheI 432
1333 CGTTACCTCGCTGTCTGGCGGTATGGGGCCACCATTTGCTG.... 1378
432 leAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThr 448
1379CAGATGCTGCTGCTGGTGAAC 1399

```

Percent Similarity: 56.916   Percent Identity: 29.478

alignment_block:
US-09-911-667A-2 x US-09-040-444-4   ..

Align seg 1/1   to: US-09-040-444-4   from: 1   to: 1882

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
:|||||:|||||   |||:   |||   |||:   |||:   |||:   |||:
473 GCCTGGAAGTGGACCTTTTTCAGTCCCTGTGTGAACTTGGCGTCTCTCTCT 522

135 rSerSerThrTrpGlyAsnIleSerAspGlnTrpGlyArgGlyThrG 152
:||||   :||   |||:   |||:   |||:   |||:   |||:   |||:
523 GGGCTCCCTGGTTGGGTTACATTGCAGACAGGTTTGGCCGTAAGCTCT 572

152 lLeuLysIleSerValLeuTrpThrLeuTrpTyrrGlyIleLeuSerAla 168
|||   |||   |||:   |||   |||   |||:   |||:   |||:   |||
573 GTCTCTGGTGACCAGCGTGGTCACATCTGTGTCGCGGTGTGCTAACAGCG 622

169 PheAlaProValTyrrTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||   |||:   |||:   |||:   |||:   |||:   |||:   |||:
623 GTGGCCCCAGACTATFACATCCATGTGTCTTTTCCCTGCTCGAGGCAAT 672

185 e...GlyIleGlyValProGlnSerValThrLeuTrpAlaGluPheL 201
:||||   :|||:   |||:   |||:   |||:   |||:   |||:   |||
673 GGTGACGAAGGCAGCTGGGTCTCGGGCTATACCTTGATCACAGAGTTTG 722

201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
:||||   |||   |||   |||:   |||:   |||:   |||:   |||
723 TCGGCTCTGGCTACAGAGAACGACGGCCATTTGTACCAGATGGCCTTTC 772

218 AlaIleGlyThrValPheGluValValLeuAlaValPheValMetProSe 234
:|||||:|||||   |||   :|||:   |||:   |||:   |||:   |||:
773 ACAGTGGGCGTAGTGGGCTTCCCGGGTGGCC...TATGCCATTCACAGA 819

234 rLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
:|||||:|||||   |||   :|||:   |||:   |||:   |||:   |||
820 C.....TGGCGCTGGCTCCAGCTAGCTGTGCTCCCTGCTACCTCTCTCT 863

251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
||   |||   :|||:   |||:   |||:   |||:   |||:   |||:
864 TCCTGCTGTATTACTGGTTTGTCCAGAAATCCCCCGGTGGCTGTGTCTCC 913

268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGl 284
:||||   :|||:   |||:   |||:   |||:   |||:   |||:   |||
914 CAGAAGAGAACCACGCGAGCTGTCCAGGATAATGGAGCAAAATTCACAGAA 963

284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
|||||   |||   |||:   |||:   |||:   |||:   |||:   |||
964 GAACGGGAAGGTGCCCTCTGCTGACTGAAGATGCTGCTCTTGAGGAGG 1013

301 spArgGlyLysMetArg.....AspLeuPhe...ThrPro 311
||   :|||:   |||   |||:   |||:   |||:   |||:   |||
1014 ATCCCTCAGAAAGCGAAGTCTTCGTTTCCGACCTGTTCCGCACCTCCC 1063

312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
:||||:|||||   |||:   |||:   |||:   |||:   |||:   |||:
1064 AACCTGAGGAAGCACACCGCTCATCTGTATCTATGTTCTCTCTGTGTC 1113

328 aPheSerTyrrTyrrGlyValLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
|||   |||   |||:   |||:   |||:   |||:   |||:   |||
1114 TGTGCTGTACCAGGTCTCATCATG.....TACTGTGACTTCTTTTA 1182

345 lyAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
:||||   :|||:   |||:   |||:   |||:   |||:   |||:   |||
1139 .....CACGTGGGAGCCACAGGGGGCAACCTC 1165

362 SerLeuAlaCysGluTyrrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
|||:   |||:   |||:   |||:   |||:   |||:   |||:   |||
1166 .....TACTGTGACTTCTTTTA 1182

378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleLeu 395
:||||:|||||   |||:   |||:   |||:   |||:   |||:   |||:

```

```

; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEO ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-501-572-6

alignment_scores:
    Quality: 426.00      Length: 429
    Ratio: 1.760        Gaps: 14
    Percent Similarity: 56.410      Percent Identity: 29.371

alignment_block:
US-09-911-667A-2 x US-08-501-572-6 ..

Align seg 1/1 to: US-08-501-572-6 from: 1 to: 1896

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||: ||| ||| |||||:|||||:
580 TCCTGGATGTTGGACCTATTCCAGTCATCAGTCGAATGAGGATTCCTTAT 629

135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTrpGlyArgLysThrG 152
:::||||| ||| |||||:|||||:|||||:|||||:
630 TGGCTCTATGAGTATCGGCTACATCAGCACAGGTTGGCCGTAAGCTCT 679

152 IyLeuLysIleSerValLeuTrpThrLeuTrpTrpGlyIleLeuSerAla 168
||| |||:|||||: ||| |||||:|||||:
680 GCCTCCTTAACACAGTCCTCATAAATGCTCAGCTGAGTTCTCTATGGCC 729

169 PheAlaProValTrpSerTrpIleLeuValLeuArg.....GlyLe 182
|||||:|||||:|||||:|||||:|||||:|||||:
730 ATTTCCTCCCAACCTATACGTGGATGTTAAATTTTTCGCTTAATCCAAGGACT 779

182 uVal.....GlyPheGlyIleGlyValProGlnSerValThrL 196
||||| |||||: |||||: ||| |||: |||
780 GGTACGCAAGACAGCTGGTTAATAGGCTACATC.....C 814

196 euTrpAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
|| ||:|||||: ||| ||| |||: |||: |||:
815 TGATTACAGAAATTTGTTGGGGGAGAGATATCGGAGAACAGCTGGGGGATTTT 864

213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla.. 228
:::||||| |||||: |||||: ||| |||: |||
865 TACCAAGTTGCCTATACAGTTGGCTC.....CTGGTGCTAGCTGG 905

229 ...ValPheValMetProSerLeuGlyTrpArgTrpLeuLeuIleLeuS 244
|| ||:|||||: ||| ||| |||: |||: |||:
906 GGTGGCTTACGCACATTCCTCAC.....TGGAGGTGGTTGCAGTTCACAG 949

244 erAlaValProLeuLeuPheAlaValLeuCysPheTrpLeuProGlu 260
|||||:|||||: |||||: ||| |||: ||| |||: |||
950 TTGCTCTGCCCAACTTCTTCTTCTGCTCTATTACTGGTGCCATACCTGAG 999

261 SerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaTh 277
||| ||||: ||| ||| |||||: |||||: |||: |||:
1000 TCTCCAGGCTGGCTGATCTCCCAAGATAAGAATGCTGAAGCCATGAGAAT 1049

277 rLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysL 294
|||||:|||||: |||||: ||| |||: ||| |||: |||
1050 CATTAAGCATCGCAAGAAAATGGAAATCTCTACCCGCCCTCCCTTC 1099

294 euIleIleSerArgGlnGluAspArgGlyLys.....Met 305
::: |||: |||||: |||||: ||| |||: |||
1100 AGCGGCTGAGACTTGAAGAGGAACTGGCAAGAAATTTGAACCTTCATTT 1149

```

```

306 ArgAspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTr 321
      |||||
1150 CTTGCTTGGTCAGAACCTCCTCAGATAAGAAACATACTATGATATTGAT 1199
      :
321 pPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuTr 338
      :
1200 GTACAACTGGTTTCAGAGCTCTGTCTCTACACAGGCCTCATCATG... 1245
      :
1246 .....CACATGGCGCTTGCAGGTGAC..... 1266
      :
355 LysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGl 371
1267 .....AA 1268
      :
371 uAspTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValL 388
      :
1269 TATCTACCTGGATTCTTCTACTCTGCCCTGGTTGAATTCGCCAGCTGCCT 1318
388 euValThrLeuTrpIleLeuAspArgLeuGlyArgLysLysThrMetAla 404
      :
1319 TCATGATCATCTCAATATTCGACCAGCATGGAGCGGTTCACCTTCGGGCT 1368
405 LeuCys...PheValIlePheSerPheCysSerLeuLeuLeuPheIle.. 419
1369 GCATCAAAATATGGTTGCAGGGCAGCGCTGT..GGCTCACTTTTATATACC 1418
420 .....CysValGlyArgAsnV 425
1419 TGGTGATCTACAATGGCTAAAAATATTATTCATCTCATGCTTGGGAAGATGG 1468
425 aLeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPhe 441
1469 GGATCACAAATG.....GCCTAT 1485
442 GlnAlaAlaTyrValTyrThrProGluValTyrProThrAlaThrArgAl 458
1486 GAGATAGTCTGCTGGTCAATGCTCAGCGTGTACCCACATTCATTAGGAA 1535
458 aLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuLeuTr 475
1536 TCTTGGCGTCCACATCTGTCTCCTCAATGTGCACATTTGGTGGCATCATCA 1585
475 hrProPheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeu 491
1586 CGGCATCTCTGGTCTACCGGCTCACTAAGATCTGGCTTGAGCTCCCGCTC 1635
492 AlaValTyrSerGlyCysCysLeuLeuAlaLeuAlaSerCysPheLe 508
1636 ATGGTGTTCGGCGCTACTTGGCTTGGTTGGTGGAGGCTCTGGTGTGTGCT 1685
508 uProIleGluThrLysGlyGlyGlyGluGlnGluSer 520
1686 TCCA...GAACATAAGGAAAGCTTTGGCTGAGACC 1719

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-040-444-6

seq_documentation_block:
; Sequence 6, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundenman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and
; TITLE OF INVENTION: DNA sequences encoding it and their use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, I
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington

```

```

; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P.
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-040-444-6

alignment_scores:
  Quality: 426.00      Length: 429
  Ratio: 1.760        Gaps: 14
  Percent Similarity: 56.410      Percent Identity: 29.371

alignment_block:
US-09-911-667a-2 x US-09-040-444-6 ..
Align seg 1/1 to: US-09-040-444-6 from: 1 to: 1896

119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 TCCTGGATGTGGACCTATCCAGTCATCAGTGAATGTAGGATCTTTAT 629
135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 TGGCTCTAGTAGTACGCGTACATAGCAGACAGGTTTGGCGTAAGCTCT 679
152 LyLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 GCCTCTAACTACAGTCTCTAATAATGCTGCAGCTGGAGTTCTCATGCC 729
169 PheAlaProValTyrSerTrpIleLeuValLeuArg.....GlyLe 182
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 ATTTCCCAACCTATAGTGGATGTTAATTTTCGTTAATCCCAAGGACT 779
182 uVal.....GlyPheGlyIleGlyGlyValProGlnSerValThrL 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 GGTGAGCAACAGCGCTGGTTAATAGGCTACATC.....C 814
196 euTyrAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 TGATTACAGAAATTTGTTGGGGGAGATATCGAGAACAGTGGGATTTT 864
213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAla.. 228
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
865 TACCAAGTTGCCTATACAGTTGGGCTC.....CTGGTGTAGCTGG 905
229 ...ValPheValMetProSerLeuGlyTyrArgTrpLeuLeuIleLeuS 244
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 GGTGCTTACGCATCTCTAC.....TGGAGGTGGTTGCAGTTCACAG 949
244 erAlaValProLeuLeuPheAlaValLeuCysPheTrpLeuProGlu 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

950 TTGCTCTGCCCAACTTCTTCTTCTCTATTACTGGTGCCATACCTGAG 999
261 SerAlaArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaTh 277
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1000 TCTCCAGGTGGTGTATCTCCCAAGTAAGAATGCTGAAGCATGAGAAT 1049
277 rLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysL 294
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1050 CATTAAGCACATCGCAAGAAATAATGAAATCTCTACCCGCCCTCCCTTC 1099
294 euIleIleSerArgGlnGluAspArgGlyLys.....Met 305
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1100 AGCGCTTGAGACTTGAAGAGAGAAATGCAAGAAATTAACCCCTTCATT 1149
306 ArgAspLeuPhe...ThrProHisPheArgTrpThrLeuLeuLeuTr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1150 CTTGACTTGGTCAGAACTCCCAAGATAAGGAAACATCTATGATATTGAT 1199
321 pPheIleTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeu 338
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1200 GTACAACCTGGTTTCAGAGCTCTGTCTTACCAGGCGCTCATCATG... 1245
338 hrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLys 354
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1246 .....CACATGGCGCTTTCAGAGTGAC..... 1266
355 LysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluG1 371
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1267 .....AA 1268
371 uAspTyrMetAspLeuLeuTrpThrLeuSerGluPheProGlyValI. 388
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1269 TATCTAGTTGGATTCTTCTACTCTGCGCTGGTTGAATTCAGCTGCCT 1318
388 euValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAla 404
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1319 TCATGATCATCTCATTCAGCGCATCGGACCGCTTACCCTTGGGGCT 1368
405 LeuCys...PheValIlePheSerPheCysSerLeuLeuLeuPheIle... 419
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1369 GCATCAAAATATGGTTGCGAGGCGAGCTGTCTGCGCTCAGTTTATACC 1418
420 .....CysValGlyArgAsnV 425
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1419 TGGTGATCTACAATGGCTAAATAATATTATCTCATGCTTGGGAAGAATGG 1468
425 alLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPhe 441
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1469 GGATCACAAATG.....GCCTAT 1485
442 GlnAlaAlaTyrValTyrThrProGluValTyrProThrAlaThrArgAl 458
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1486 GAGATAGTCTGCTGGTCAATGCTGAGCTGTACCCACACATTCATTAGGAA 1535
458 alLeuGlyLeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleT 475
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1536 TCTTGGCGTCCACATCTGTTCTCAATGTGTGACATTTGTCGCATCATCA 1585
475 hrPropheIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeu 491
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1586 CGCCATCTCTGGTCTACCGGCTCACTAACATCTGGCTTGAGCTCCCGCTG 1635
492 AlaValTyrSerGlyCysLeuLeuAlaAlaLeuAlaSerCysPheLe 508
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1636 ATGTTTTTCGGCGTACTTGGCTTGGTGTGGAGTCTGGTGTGCTTGTGCT 1685
508 uProIleGluThrLysGlyGlyLeuGlnGluSer 520
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1686 TCCA...GAAACTAAGGAAAGCTTTGCCTGAGACC 1719
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-501-572-5
seq_documentation_block:

```

```

: Sequence 5, Application US/08501572
: Patent No. 6063623
: GENERAL INFORMATION:
: APPLICANT: Koepsell, Hermann
: APPLICANT: Grundeman, Dirk
: APPLICANT: Gorboulev, Valentin
: TITLE OF INVENTION: Transport protein which effects the
: TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
: TITLE OF INVENTION: DNA sequences encoding it and their use.
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/501,572
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Toohey, Kimberlin M
: REGISTRATION NUMBER: 35,391
: REFERENCE/DOCKET NUMBER: 02481.1453-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)408-4000
: TELEFAX: (202)408-4400
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1885 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-501-572-5

```

```

alignment_scores:
  Quality: 425.50      Length: 421
  Ratio: 1.766         Gaps: 11
  Percent Similarity: 57.245  Percent Identity: 30.166

alignment_block:
US-09-911-667A-2 x US-08-501-572-5

Align seg 1/1 to: US-08-501-572-5 from: 1 to: 1885

119 SerTrpGlnValAlaLeuLeuThrSerValAlpHeValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 TCCTGGAAGCTGGACCTCTTTCAGTCTGTTGAATGCGGCTCTTCTT 554
135 rSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
555 TGCTCTCTCGGTGGCTGCTTTCAGACAGGTTCGCGGCTGCTGCTGCTG 604
152 LyLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 GTCTCTGGGAAGCTGCTGCTCAACGCGGCTGCGGCGTCTGCTGCTG 654
169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 TTCTCGGCCAACATACATGCTCATGCTCTTCCGCTGCTGCTGCTGCTG 704
185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheL 201
: : : : : : : : : : : : : : : : : : : : : : : : : :
705 GGTGACGACGAGCACTGGATGCTGCTACACCCCTAATCACAGAAATTG 754

```

```

201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
: : : : : : : : : : : : : : : : : : : : : : : : : :
755 TTGGCTCGGGCTCCAGAGACGGTGGCGATCATGTATACAGATGGCCTTC 804
218 AlalleGlyThrValPheGluValValLeuAlaValPheValMetProse 234
: : : : : : : : : : : : : : : : : : : : : : : : : :
805 ACGGTGGGGCTGGTGGCGCTTACCGGGCTGGCC...TAGCCCTGCTCA 851
234 rLeuGlyTyrArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
852 C.....TGGCGCTGGCTGCAGCTGCAGTCTCCCTGCCACCTTCTCT 895
251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
|| : : : : : : : : : : : : : : : : : : : : : : : : :
896 TCCTGCTCTACTACTGTGTGTCGGAGTCCCTCGGTGGTGTATCA 945
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG 284
: : : : : : : : : : : : : : : : : : : : : : : : : :
946 CAAAAAGAAACACTGACGCAATAAGACATAATGACACCATCGCTCAAAA 995
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 GAATGGGAAGTTGCTCTCTGCTGATTTAAAGATGCTTCCCTCGAAGAG 1045
301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
|| : : : : : : : : : : : : : : : : : : : : : : : : :
1046 ATGTCACCGAAAAAGCTGAGCCCTTCATTTGACAGACCTGTCCCGACGCG 1095
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnA 328
: : : : : : : : : : : : : : : : : : : : : : : : : :
1096 CGCCTGAGGAAGCGCACCTTCATCTGATGTACCTGTGTGGTTACGGACT 1145
328 apheSerTyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
: : : : : : : : : : : : : : : : : : : : : : : : : :
1146 TGTGCTCTATCAGGGGCTCATCTCTG.....CACATGG 1177
345 LyAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
|| : : : : : : : : : : : : : : : : : : : : : : : : :
1178 GCGCCACGACGCGG..... 1191
362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
: : : : : : : : : : : : : : : : : : : : : : : : : :
1192 .....AACCTCTACCTGGATTTCCCTTTA 1214
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleIleA 395
: : : : : : : : : : : : : : : : : : : : : : : : : :
1215 CTCCTGCTCTGTGCAAAATCCCGGGGCTTCATAGCCCTCATCACCATTG 1264
395 spArgLeuGlyArgLysLysThrMetAlaLeu..... 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1265 ACCGCTGGGCGCGCATCTACCCCATGCGCTGCAAAATTTGTTGGCGGG 1314
406 .....CysPheValIlePheSerPheCysSerLeu.....Le 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 GCAGCTGCTGCTGCTGCTATTTTATCTACCTGACCTGCACCTGTTAAACAT 1364
416 uLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleA 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1365 CATAATCATGTGTGGCGCAATGGGAATCACCATTGCAATA..... 1407
433 laArgAlaPheIleSerGlyPheGlnAlaAlaTyrValTyrThrPro 449
|| : : : : : : : : : : : : : : : : : : : : : : : : :
1408 .....CAAATGATCTGCTGCTGGTGAATGCT 1431
450 GluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerG 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1432 GAGCTGTACCCACATTCGTCAGGAACCTCAGATGATGTGTCTCTC 1481
466 yMetAlaArgValGlyAlaLeuIleThrPropheIleAlaGlnValMetL 483
: : : : : : : : : : : : : : : : : : : : : : : : : :
1482 CCTGTGTGACATAGGTGGGATAATACACCCCTTCATAGTCTTCAGGCTGA 1531
483 euGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeu 499

```

```
1532 GGGAGGCTGTGCAAGCCTTGGCCCTCATTTTGTGTGGCGTGTGGGCGTG 1581
500 LeuAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyG1 516
1592 CTTCGGCGGGAGTGGACGCTACTTCTTCCA...GAGACCAAGGGGAGCG 1628
516 yLeuGlnGluSer 520
1629 TTTCGACAGAC 1641
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-040-444-5
seq_documentation_block:
; Sequence 5, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein Which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-5
```

```
alignment_scores:
Quality: 425.50 Length: 421
Ratio: 1.766 Gaps: 11
Percent Similarity: 57.245 Percent Identity: 30.166
```

alignment_block:

US-09-911-667A-2 x US-09-040-444-5 ..

Align seg 1/1 to: US-09-040-444-5 from: 1 to: 1885

```
119 SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSe 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 TCCTGGAGCTGGACCTCTTTCAGTCCTGTGTGAATGCGGCTCTCTCT 554
135 rSerSerThrLeuTrpGlnValAlaLeuThrSerValValPheValGly 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
555 TGGCTCTCTCGGTGTGGCTACTTTTGCACAGAGTTTGGCCGTAAAGCTGT 604
152 lyLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAla 168
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 GTCTCTCGGAACTGTGTGTCAACGGCGTGTGGGCGTGTCTCATGGCC 654
169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPh 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
655 TTCTCGCCCAACTACATGCTCCTCTTCCGCTGTGTCAGGGCGCT 704
185 e...GlyIleGlyValProGlnSerValThrLeuTyrAlaGluPheLe 201
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
705 GGTCAAGCAAGGCAACTGGATGGCTGTGTACACCTCAATACAGAAATTG 754
201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 TTGGCTCGGGCTCCAGAAAGACGGTGGCGATCATACCATGATGGCCTTC 804
218 AlaIleGlyThrPheGluValValLeuAlaValPheValMetProSe 234
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
805 ACGTGGGGCTGGTGGCGCTTACCGGGCTGGCC...TAGCCCTGCCTCA 851
234 rLeuGlyTrpArgTrpLeuIleLeuSerAlaValProLeuLeuLeuP 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
852 C.....TGGCGCTGGCTGCAGCTGGCAGTCTCCCTGCCACCTTCTCT 895
251 heAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspValLeu 267
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
896 TCCTGCTCTACTACTGGTGTGTGGCGGAGTCCCTCGGTGGTGTATCA 945
268 SerGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrG1 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 CAAAAAGAAACACTGAAGCAATAAAGATAATGACACACATCGCTCAAAA 995
284 uAsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGluA 301
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
996 GAATGGGAAGTTGCCTCTGCTGATTTAAAGATGCTTCCCTCGAAGAGG 1045
301 sp.....ArgGlyLysMetArgAspLeuPhe...ThrPro 311
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1046 ATGTCACCCGAAAGCTGAGCCCTTCATTTGCAGACCTGTTCCGCACGCG 1095
312 HisPheArgTrpThrThrLeuLeuLeuTrpPheIleTrpPheSerAsnAl 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1096 CCGCTGAGGAAGCGACCTTCATCTCTGATGTACCTGTGGTTCCAGGACT 1145
328 apheSerTyrTyrGlyLeuValLeuThrThrGluLeuPheGlnAlaG 345
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1146 TGTGCTCTATCAGGGCGCTATCTG.....CACATGG 1177
345 lyAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLysCys 361
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1178 GCGCCACCGAGCGG..... 1191
362 SerLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuTr 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1192 .....AACCTCTACCTGGATTTCTCTTA 1214
378 pThrThrLeuSerGluPheProGlyValLeuValThrLeuTrpIleAla 395
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1215 CTCGGCTGTGGTCAAAATCCCGGGGCGCTTCATAGCCCTCATCACCATTG 1264
395 spArgLeuGlyArgLysLysThrMetAlaLeu..... 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1265 ACCGGTGGCGCGCATCTACCCCATGGCGGTGTCAAAATTTGTTGGCGGG 1314
406 .....CysPheValIlePheSerPheCysSerLeu.....Le 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 CGAGCGCTGCCTCGTCATTTTATCTACCTGACCTGACCTGGTTAAACAT 1364
416 uLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleA 433
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1365 CATATCATGTGTGTGGCCGAATGGGAATCACCATTGCAATA..... 1407
```


433 IaArgAlaPheIleSerGlyClyPheGlnAlaLaTy rValThyThrPro 449
||| : : : :
1408CAAAATGATCTGCCTCGAATGCT 1431

450 GluValTy rProThrAlaThrArgAlaLeuGlyLeucl yThrCy sSerGI 466
||||: ||||| : : ||| ||| : : ||||| :
1432 GAGCTGTACCCACACTTGGTCAGGAACCTCAGAGTGATGGTGTGTTCCCTC 1481

466 YMetAla rGValGlyAlaLeuIIeThrProPheI leAlaGlnValMetL 483
:
1482 CCTGTGTGCATAGTAGGGATAATCACCCCCTCATAGTCTTCAGGCTGA 1531

483 euCl uSerSerValTy rLeuThrLeuAlaValTy rSerGI cy sCysLeu 499
||| ||| ||| : : : : : : : :
1532 GGGAGGCTGCCAGGCTTGCCTCTATTGTTGGGTGTGGGCCTG 1581

500 LeuAlaAlaLeuAlaSerCy sPheLeuProfl eCl uThrly sGlyGI yGI 516
||||| :
1582 CRTTGGCGGGAGTGACCGCTACTTCTCCA...GAGACCAGGGGGACGC 1628

516 yLeuGlnGluSer 520
||| ||| : : : :
1629 TTTCGCAGAGACC 1641

seq_name: /cqn2_6/ptodata/2/ina/5B_COMB.seq:US-08-647-397-1

```

seq_documentation_block:
: Sequence 1, Application US/08647397
: Patent No. 5972702
: GENERAL INFORMATION:
: APPLICANT: Beier, David R.
: APPLICANT: Brady, Kevin P.
: TITLE OF INVENTION: OSTEOCLAST TRANSPORTER
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/647,397
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Gates, Edward R.
: REGISTRATION NUMBER: 31,616
: REFERENCE/DOCKET NUMBER: B08017048
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-720-3500
: TELEFAX: 617-720-2441
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2102 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 120..1733

```

[illegible]

223 PheGluValValLeuAlaValPheValMetProSerLeuGly..... 236
786TTCAATTCCTGCGCTGGCCCTGACCTATGCCGT 814
237TrpArgTrpLeuLeuLeuSerAlaValProLeuLeuLeuP 251
815 TCCTCAGTGGCGTGTACAGTTGTCGCTGCTGCTGCTTCTTCATCT 864
251 heAlaValLeuCysPheTrpLeuProGluSerAlaAlaArgTyrAspValLeu 267
865 TCTCCTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914
268 SerGlyAsnGlnGluLysAlaAlaAlaThrLeuLysArgIleAlaThr... 283
915 TCTGAAAAATTCACAGAGCTCTGAAGACACTCAACAGTGTGGGTACTCT 964
284GluAsnGlyAlaProMetProLeuGlyLysLeuLeuI 296
965 CAACGGCAAGGAAGGAAGGAAAGCTCACTGTGGAGGAGTGAAGT 1014
296 leSerArgGlnGluAsp.....ArgGlyLysMetArg 306
1015 TCAACTTCGAGGAAGGACATCACTCAACCAAGGTCAAAATATGGCTTATCT 1064
307 AspLeuPhe...ThrProHisPheArgTrpThrThrLeuLeuLeuTrpPh 322
1065 GACTTGTTCGAGTGTCCATCTCGCGCGTGTGACCTTCTGCTCTCTCT 1114
322 eileTrpPheSerAsnAlaPheSerTyrTyrGlyLeuValLeuLeuThrT 339
1115 GGCCTGCTTGTCTACTGCTGCTGCTACTACAGTGTGCTATGGGTAG 1164
339 hrGluLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLys 355
1165 AAGAATT.....GGAGTCAACATC..... 1184
356 AlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAs 372
1185TACATA..... 1190
372 pTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuV 389
1191CTCAGATCATCTTCGGTGGGT...GACATTCGCCCAAGTTCA 1233
389 alThrLeuTrpIleAlaAspArgLeuGlyArgLysLysThrMetAlaLeu 405
1234 TCACAATCCTCTCCATAAGTTATCTGGCGCGCATCACTCAGGCGTTC 1283
406 CysPheV:lIlePheSerPheCysSerLeu...LeuLeuPheIleCysVa 421
1284 CTCCTGATCCGCGAGGAGTGGCCATCTGCGCCTCATCTTGTGTCTTC 1333
421 lGlyArgAsnValLeu...ThrLeuLeuLeuPheIleAlaArgAlaPheI 437
1334 AGAAATCAGCTCTTGAGACAGCACTGGCTGTATTGGGAAGGATGCC 1383
437 leSerGlyGlyPheGlnAlaAlaTyrValTyrThrProGluValTyrPro 453
1384 TGTCTGGCTCTTCAGCTGGCTCTCTCTACACAAGTACAGCTACACCT 1433
454 ThrAlaThrArgAlaLeuLeuGlyLeuGlyThrCysSerGlyMetAlaArgVa 470
1434 ACAGTCTCAGCAACAGGTATGGGTATCAGTAACATATATGGGTCCAGT 1483
470 lGlyAlaLeuIleThrProPheIle.....AlaGlnValM 482
1484 GGAAGTATGATACCCCACTGTTGAAATCACGGGAGAACTGCAGCCCT 1533
482 etLeuGluSerSerValTyrLeuLeuAlaValTyrSerGlyCysCys 498
1534 TCATCCCAATGCTATCTTTGGACCATGACTCTACTGGGAGGC..... 1577
499 LeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyI 515

1578AGTGTGCTTCTTCTGCTTGAGACCTCAATCG 1612
515 yGlyLeuGlnGluSerSerHis.....ArgGluTrpGlyGlnGlu 528
1613 GCCCTTACCAGAACTATCGAGGACATACAAGACTGGTACCAGCAA 1658
seq_name: /cyn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
alignment_scores:
Quality: 230.00 Length: 476
Ratio: 0.975 Gaps: 14
Percent Similarity: 49.580 Percent Identity: 22.479
alignment_block:
US-09-911-667A-2 x US-09-103-840A-1 ..
Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529
86 LeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGluMetMe 102
3717266GTCGCGCTCTCTACGGGTACGACCTTTCGCCCATCGCGGTGCTGCT 3717315
102 tIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLeuProS 119
3717316GTCCTCAGCGAG.....CAATTCGAACCTCACCA 3717344
119 erTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMetSer 135
3717345CTCAGAAACAGGAGTGTGTACCAACACAGCGGTGCTCGGCAGATCGCC 3717394
136 SerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrG 152
3717395GGGCGCTTGGCGCGCATCTCGCCCAACCGCATCGACCAAGAAATC 3717444
152 yLeuLysIleSerValLeuTrpThrLeuTyrTyrGlyIleLeuSerAlaP 169
3717445GCTGCTCATCTCGCGCGCTACGCGAGTGTGCGCTGCTCGCGCGGA 3717494
169 heAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPhe 185
3717495CCTCGGTGCTGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTG 3717544
186 GlyIleGlyValProGlnSerValThrLeuTyr...AlaGluPheLe 201
3717545ACAATCGCCTGTGCGGTGCTGCGGTGCTGCTGCTGCTGCTGCGCG 3717594
201 uProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrpA 218
3717595GCCCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3717644

```
218 laileGlyThrValPheGluValValLeuAlaValPheValMetProSer 234
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717645TTAGCGGATCGTCTCGTGGTACCTGGTCCGCTACCTGTTGGCCGATCG 3717694

235 LeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeuPh 251
||||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717695CACGCGTGGCGGATCTTGGGTCGCGCGCGCGCGCGGATCGTCT 3717744

251 eAlaValLeuCysPheTrpLeuProGluSerAlaArgTrpAspValLeu 268
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717745GTTGTCGTTGTTGGCGCATGCCGATACCCCGCTGGTATCTGCTCA 3717794

268 erGlyAsnGlnGluLysAlaIleAlaThrLeuLysArgIleAlaThrGlu 284
: : : : : : : : : : : : : : : : : : : : : : : : : :
3717795AGGCGCGGATCGCCGACGCGGTAGCGCTGCGCGGATCGACCGGAG 3717844

285 AsnGlyAlaProMetProLeuGlyLysLeuIleIleSerArgGlnGlu 301
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717845CGCGATCGATCGCGAGCTTCCTCGTCCAGATCACCGGGATCAACGCG 3717894

301 pArgGlyLysMetArgAspLeuPheThrProHisPheArgTrpThrL 318
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717895CGCGCGGATCGCGAAATGTCGGCGCGCTATCTGCGGGCCACGCG 3717944

318 euLeuLeuTrpPheIleTrpPhe.....SerAsnAla 328
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717945TGTTCGTCATCGCGCTCGCTTCCTCGTCCAGATCACCGGGATCAACGCG 3717994

329 PheSerTyrrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG 345
||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3717995ATCATCTACTAC.....AGTCCGCGACATTTTCGCGCCCAT 3718029

345 yAspValCysGlyIleSerSerArgLysLysAlaValGluAlaLys 362
: : : : : : : : : : : : : : : : : : : : : : : : : :
3718030GGCTTCGCGGCG..... 3718042

362 erLeuAlaCysGluTyrrLeuSerGluGluAspTrpMetAspLeuLeu... 377
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718043.....TATTCGCGATGCTTGGC 3718060

378 .....TrpThrThrLeuSerGluPheProGlyValLeuValThrIe 391
: : : : : : : : : : : : : : : : : : : : : : : : : :
3718061CTGCGCGGATCGTCAAGTCCCGCTTGGCGCGGCTGTGCTCGCT 3718110

391 utrpIleIleAspArgLeuGlyArgLysLysThrMet..... 403
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718111GTTTGTGTCGATCGCTGGCGCGCTCGCCGATCTGTTGTCGCGCATCG 3718160

404 .....AlaLeuCysPheValIlePheSer..... 411
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718161CGACGATGATCACCGCAGATCGCTGCTGATCACCGCTATTCGCCAACGAC 3718210

412 .....PheCysSerLeuLeuLeuPh 418
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718211TCGATGTTGGCAGCGGGCTGTGTTGGGTTTCGCGCGGCTGCTGTT 3718260

418 eIleCysValGlyArgAsnValLeuThrLeuLeuPheIleAlaArgA 435
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718261CATC....ATCGGGTTCACTTC..... 3718279

435 laPheIleSerGlyGlyPhe...GlnAlaAlaTyrrValTyrrProGlu 450
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
3718280.....GGATTCGGCTCGCTGCTGGGTGTGACGCCGCGGAG 3718315

451 ValTyrrProThrAlaThrArgAlaLeuGlyLeuGly...ThrCysSerG 466
: : : : : : : : : : : : : : : : : : : : : : : : : :
3718316AGCTTCCCTCCCGCTGGGTGCGATGGGATCGAGCCCGATGCTCACCTC 3718365

466 yMetAlaArgValGlyAlaLeuIleThrProPhe.....IleAlaG 480
: : : : : : : : : : : : : : : : : : : : : : : : : :
3718366GACACTGACGGCAACGCGATGCTTGGCCGCTTCTCGCTCACCATGCTGC 3718415
```

```

732 CTTCCTGTGCTACCAACCAACGATCGGCCAGTGGGATCTGGTGTGA 781
119 ...SerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetM 134
...:|||||  |||  ::  :::::||||:
782 CCTGGGCTGCAGGTATCTCGAGCAGATCTCTTCATCTGGCTTG 831
134 etSerSerThrLeuTrpGlnPheValSerAspGlnTrpGlyArgLys 150
||||:  ::  |||  ::|||:|||||:||||:
832 CCTCGGCTACCTGTTCTGGGTACCCGACAGACAGATTGGCCGTCG 881
151 ThrGlyLeuLysIleSerValLeuThrThrLeuTyrTyr..... 163
||||:  |||||  |||||  |||||
882 ...GGGATT.....GTGCTGCTACCTTGGGCTGGTGGCCCTG 919
164 GlyIleLeuSerAlaPheAlaProValTyrSerTrpPheLeuValLeuA 180
||||:  |||  |||  ::  :::::||||:
920 TGGAGTAGGAGGGCTGCTGAGGCTCTCCACAGGCTCATGGCCCTCC 969
180 rgGlyLeuValGlyPheGlyIleGlyValProGlnSerValThrLeu 196
|||  |||:|||||  :::::|||||  ::|||  |||
970 GATTCCCTTGGGCTTCTGCTGGGTTGACCTGGGTGTCTACCTG 1019
197 Tyr...AlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLe 212
||||:  |||  |||  ::  :::::||||:
1020 ATGCGCTGGAGCTGCGACCAACCAAGGCTTCGGGTGGCCCTGGC 1069
212 uIleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAlav 229
||||:  |||  |||  ::  :::::||||:
1070 AGGGGAGTTG.....GTGGGGTGGAGGCACTTCTGTCCTGG 1110
229 alPheValMetProSerLeuGlyTrpArgTrpLeuLeuLeuSerAla 245
:::  :::::  |||  |||||:||||:  ::  :::
1111 GCCTGGCCCTGCTCTAAGGATTGGGATTCCTACAGCAATGATCACC 1160
246 ValProLeuLeuPheAlaValLeuCysPheTrpLeuPro..... 259
||||  ::|||  |||  |||  ::  |||
1161 GCTCCCTGCATCCTCTTC.....CTGTTTATGGCTGGCTGGTGTGT 1204
260 ....GluSerAlaArgTrpAspValLeuSerGlyAsnGlnGluLysAlaI 275
|||||  |||||  |||  ::  :::::||||:
1205 CTGGAGTCCCGAGCTGGGTGATGTAAGCGGCGAGATTGAGGAGCTC 1254
275 leAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetProLeu 291
:::  |||  |||  ::  |||  :::::||||:
1255 AGTCTGTGCTGAGGATCTGCTGAGCGAAAC.....CGGCCCAT 1295
292 GlyLysLeuIleIleSerArgGlnGluAspArgGlyLysMetArgAspLe 308
||||:  :::::  |||  |||  ::  |||  :::::||||:
1296 GGGCAGATGCTG.....GGGAGGAGGCCAGGAGGCCCTGCGAGGACCT 1339
308 u.....PheThrProHisPheArgTrpThrL 318
|  |||  |||  |||  ::  |||
1340 GGAGAATACCTGCCCTCTCCCTGCAACATCTCTCTTTCTTGTCTCC 1389
318 euLeu.....LeuTrp.....PheIleTrpPheSer 326
||||  ::|||  ::|||  ::|||  ::|||
1390 TCCTCACTACCGCAACATCTGAAAAATCTGTTATCTCTGGGCTTCACC 1439
327 AsnAlaPheSerTyrGlyLeuValLeuLeuThrThrGluLeuPheG1 343
|||  |||  ::|||  ::|||  ::|||
1440 AAC...TTTATGCCCAGTCCATTCGCCACTGCTACCGAGCTGGGAGG 1486
343 nAlaGlyAsp.....ValCysGlyIleSerSerArgLysL 355
::|||:  ::|||:  ::|||:  ::|||:  ::|||:
1487 AGGAGGAGGCCATCGGACTTCTACCTGTGCTCTGCTGGCCAGCGCA 1536
355 ysAlaValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGlu 371
|||  |||  |||||  |||||  |||||
1537 CCGCA.....GCCCTGGCTGTCTTCCTG..... 1562
372 AspTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLe 388
||||  |||||
1563 .....GGGCTCAC 1570

```

```

388 uValThrLeuTrpIleIleAspArgLeuGlyArgLysLysThrMetAlaL 405
|||  |||||  |||||  |||||  |||||
1571 CGTG.....GACCATTTGGCGCGGGGATCCTCTCTTC 1605
405 euCysPheValIlePheSerPheCysSerLeuLeuLeuPheIleCysVal 421
|||  |||  ::|||  ::|||  ::|||  ::|||
1606 TCTCATGACCTTACCGCATTCCTCCCTGGTCTG. .... 1643
422 GlyArgAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSe 438
|||  |||  ::|||  ::|||  ::|||  ::|||
1644 .....CTGGGCTCTGGGATTATCTGAACGAGGCTGCCATCAC 1681
438 r.....GlyGlyPhe.....GlnAlaAlaTyrValTyrT 448
:  |||  |||  |||||  |||||  |||||
1682 CACTTCTCTCTGCTGGCTCTCTCTCCCAAGCTGCCGCATCCTCA 1731
448 hr.....ProGluValTyrProThrAlaThrArgAlaLeuGly 460
|||  |||||  |||||  |||||  |||||  |||||
1732 GCACCTCTTGTCTGCTGAGTCACTCCACACACTGTCGGGGCGGTGC 1781
461 LeuGlyThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPh 477
||||  ::|||  ::|||  ::|||  ::|||  |||
1782 CTGGCCTGATCATGGCTCTAGGGCGCTTGGAGACTGAGCGCCG... 1829
477 eIleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaValT 494
:::  :::::  ::|||  ::|||  ::|||  ::|||
1830 .GCCAGCGCTCCACATGGGCGGAGCTTCTGCGACGACGTTGGTGC 1878
494 yrSerGlyCysCysLeuLeuAlaLeuAlaSerCysPheLeuProIle 510
::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1879 TGGCGGCTGCGCCCTCTCTGCTATCTCAGCATTCGTGCTGCTGCCG... 1925
511 GluThrLys 513
|||||  |||||
1926 GAGACCAAG 1934

```

seq_name: /cgn2 6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:

```

; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
Quality: 246.50 Length: 534
Ratio: 0.970 Gaps: 19
Percent Similarity: 47.566 Percent Identity: 22.097

alignment_block:
US-09-911-667A-2 x US-09-103-840A-2/rev ..

```

[illegible][illegible]

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1

```
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRATER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES F
; TITLE OF INVENTION: TUBERCULOSIS
```

```

; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
; US-09-103-840A-1

```

```

alignment_scores:
  Quality: 246.50      Length: 534
  Ratio: 0.970         Gaps: 19
  Percent Similarity: 47.566      Percent Identity: 22.097

```

```
alignment_block:
  US-09-911-667A-2 x US-09-103-840A-1/rev ..

```

```
Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529
```

```

12 ProValVallys.....PheArgArgThrGlyCysLeuSerAlaArgse 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150410CCCCAATATGCGATCCGGTCACTCGCGCGCGCATCAGAAAGGCCACATT 2150361
25 rGluAspThrAlaSerGlyGluHisGluValGlnIleGluGlyValH 42
||| :|||:|||||:
2150360GCCCGCATCTGCCACGGAC..... 2150342
42 isValGlyLeuGluAlaValGluLeuAspGlyAlaAlaValProLys 58
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150341...ATGCGAGCGCGCGTGTTCGCGCATCTCCCGCATCTCCCGCGCGCG 2150294
59 GluPheAlaAsnProThrAspAspThrPheMetValGluAspAlaValG 75
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150293CCGTCACCGCATGCGAGC.....GTGCG 2150268
75 uAlaIleGlyPheGlyLysPheGlnTrpLysLeuSerValLeuThrGlyL 92
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150267ASCACCGCGGTGTGACCGCGCACCGAGCGAAACGCTTTATGGCATCTTCT 2150218
92 euAlaTrpMetAlaAspAlaMetGluMetIleLeuSerIleLeuAla 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150217TGGGTGACGATGCGTTCATTTCTCTCGTGGTCTGTCTAT 2150168
109 ProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnValAlaLeuLe 125
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150167GCCGACATCGAACGACCTTCCACACACCAAGACCGACGTCGCGTTTCT 2150118
125 uThrSerValValPheValGlyMetMetSerSerSerThrLeuTrpGlyA 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150117CACACCGCCACCTGGCATGCCCGCGCGGTGGGTGGTGTCTGTCTGGCG 2150068
142 snIleSerAspGlnTrpGlyArgLysThrGlyLeuLysIleSerValLeu 158
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150067TATGGCGCGACCGGTGGCGCGCGCGGTCCGCTGATGTCGACGTCTCG 2150018
159 TrpThrLeuTrpGlyIleLeuSerAlaPheAlaProValTrpSerTr 175
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2150017TTCATTCGGTGATCGATTCTGTGCGCATTCGACCCCACTTTCACCGT 2149968
175 pIleLeuValLeuArgGlyLeuValGlyPheGlyIleGlyGlyVal...P 191
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149967GCTGTGATCTGCGACTGCTCTACGGCATTTGGCATGGCGGAGAGTGG 2149918
191 roGlnSerValThrLeuTrpAlaGluPheLeuProMetLysAlaArgAla 207
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149917GGGTGGCGCGCTGTCTCCATGGAGAGGTTCAGCGCGAGCGCGTGGC 2149868
208 LysCysIleLeuLeuIleGluValPheTrpAlaIleGlyThrValPheG 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

2149867GTCTTCTCCGGGCTGTGTGACAGAGGGCTACGGCTTCGGCTATCTGTGGC 2149818
224 uValValLeuAlaValPheValMetProSerLeuGly.....TrpArgT 239
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149817CAGTGTGGCAGCGCTGTGTGATGAACCTAGGCTGTGTGGCGGT 2149768
239 rPleuLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCys 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149767GTGTGTTCGGCCTAAGCATCATCCGCTCTGATCAGCTTCATCATCCG 2149718
256 PheTrpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsnGlnG 272
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149717TACCGGTAAGGATCC.....GAAGTCTGGGAAGCGCGCAGGA 2149677
272 uLysAlaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProM 289
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149676CCGCTCACCAGACCGCAATCCGCGATGCTCTGGGTAAACCC. 2149628
289 etProLeuGlyLysLeuIleLeuSerArgGlnGluAspArgGlyLysMet 305
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149627.....GCAATTGTGCGCGG..... 2149613
306 ArgAspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrp... 321
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149612.....TTGTCTACCTAGTGTGTGTGATGACCGC 2149584
322 PheIleTrpPheSerAsnAlaPheSer.....TyrTrpGlyLeuValL 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149583CTTCAACTGGATGACCGCACCGCATCTACCCGACCTTCCTGA 2149534
336 euLeuThrThrGlu..... 340
|||||:|||||:
2149533CCGCGACTACCGCATCGTCCCGGTTGTCCAGCTCACCGCCAGGTGG 2149484
341 .....LeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLy 354
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149483ATAGTGTGTGATCTACAACTCGCGCATCATCGGC..... 2149448
354 sLysAlaValGluAlaLysCysSerLeuAlaCysGluTrpLeuSerGluG 371
2149448..... 2149448
371 luAspTyrMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyVal 387
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149447.....GGTCTGGCGTTCGGCACGTTGTCCAG..... 2149421
388 LeuValThrLeuTrpIleLeuAspArgLeuGlyArgLysLysThrMetAl 404
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149420.....CGTTACCGCGCGCTTACACCATCGT 2149395
404 aLeuCysPheVal.....IlePheSerPheCys 414
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149394TTTCTGTCCGCGCTGGGACTGCCGATGTCGCGCTGTCTACTCGC 2149345
414 erLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuLeu 430
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149344GCACCTGGCGCATGTGTGCTGGTTCG..... 2149316
431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrVal 447
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149315TTTTGTGCAAGTCTCTGACAGGGTCCCTGGGTGTGATCCCGCGCA 2149266
447 rThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrC 464
::: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149265TCTCACCGAGATGTCCGACAGCTATCCGG.....GGCGTCT 2149228
464 ysSerGlyMetAla...ArgValGlyAlaLeuIleThrProPhe..... 477
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149227ACCCCGGTAACTTACCGTTCGCTAATCTGCTGGCGCGGTTCACCTT 2149178
478 ...IleAlaGlnValMetLeuGluSerSerValTyrLeuThrLeuAlaVa 493
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2149177CCTATTCAAGAGCGTCTGCGCGCAATCCCATGGCTAC..... 2149142

```

493 ltyrserGlyCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProI 510
||||| : : : : :
2149141.....CCCTTCGCTTGGCGGACGATCGTGGCCG 2149111

510 le 510

21491107G 2149109

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-591-025-8

seq_documentation_block:
; Sequence 8, Application US/09591025
; Patent No. 6303373
; GENERAL INFORMATION:
; APPLICANT: Bogan, Jonathan S.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: Method of Measuring Plasma Membrane
; TITLE OF INVENTION: Targeting of GLUT4
; FILE REFERENCE: 0399.1210-004
; CURRENT APPLICATION NUMBER: US/09/591,025
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/154,078
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: 60/138,237
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2592
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified GLUT4 containing myc tag sequences
US-09-591-025-8

alignment_scores:
Quality: 154.50 Length: 465
Ratio: 0.709 Gaps: 23
Percent Similarity: 46.882 Percent Identity: 23.226

alignment_block:
US-09-911-667A-2 x US-09-591-025-8 ..

Align seg 1/1 to: US-09-591-025-8 from: 1 to: 2592

123 AlalaLeuThrSerValValPheValGlyMetMetSerSerThrLe 139
||||| : : : : :
559 GCCCTCTCCGTGGCCATCTTTCCGTGGCGGCATGATTTCTCTTCCT 608
139 uTrpGlyAsnIleSerAspGlnTyrGlyArgLysThrGlyLeuLysIle. 155
| ||| ||||| : : : : :
609 CATTCGTATCATCTCTCACTGGCTTGGAGGAAAGAGGCCCATGCTGCA 658
156SerValLeuThrPheLeuTyrTyrGlyIleLeuSerAla 168
: : : : :
659 ACAATGTCCTGGCGGTGCTGGGGGACCCCTCATGGCCCTGGCCAAACGCT 708
169 PheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValGlyPhe 185
||| : : : : :
709 ...GCTGCCTCTATGAATGCTCATCTCTGGACGATTCCTCATTTGGCGC 755
185 eGlyIleGlyValProGlnSerValThrLeuTyr...AlaGluPheL 201
||| : : : : :
756 CTACTCAGGGGTGACATCAGGGGCTGCTCCCATGTACGTGGGGAGATTG 805
201 euProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTrp 217
||| : : : : :
806 CTCCCACTCACTGGCGGC..... 825
218 AlaIleGlyThrVal.....PheGluValValLeuAlaValPheValMe 232
||||| : : : : :
1594 AACTGGACGAGCAACTTCATCATTTGGC...ATGGGTTTCCAGTATGTTGC 1640


```
303 .....GlyLysMetArgAspLeuPheThrProHisPheArgTrpThrT 317
      |||... |||...|||
19730 GGGCTGGGATCGCGCGTTCGCTG.....A 19705
      |||... |||...|||
317 hrLeuLeuLeuTrpPheLeTrpPheSerAsnAlaPheSerTyrTyrGly 333
      |||... |||...|||
19704 CGCTGTC.....GCCGTCTCGTGTATCGGC 19679
      |||... |||...|||
334 LeuValLeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyI 350
      |||... |||...|||
19678 CTTGGTCTGCTTCGTG.....CT 19662
      |||... |||...|||
350 eSerSerArgLysAlaValGluAlaLysCysSerLeuAlaCysGluT 367
      |||... |||...|||
19661 GCCCAACGGCGCGCGCTGGAGCG..... 19634
      |||... |||...|||
367 yrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrLeuSerGlu 383
      |||... |||...|||
19634 ..... 19634
384 PheProGlyValLeuValThrLeuTrpIleLeuAspArgLeuGlyArgly 400
      |||... |||...|||
19633 .....ATGGTCCCGTACGGCTGTTCGGCGCGGTGGCATCAC 19596
      |||... |||...|||
400 slYThrMetAlaLeuCysPheValIle.....P 410
      |||... |||...|||
19595 GATGACACCGCGCTCAACTTCAACATCGCGCTCGCATCTTCGGTACGG 19546
      |||... |||...|||
410 heSerPheCysSerLeuLeuLeuPheIleCysValGlyArgAsnVal... 425
      |||... |||...|||
19545 TCTCCACCTGCGCTGTTCCTGCGAGCTGGTGCAGGGCGCGAGTCCACC 19496
      |||... |||...|||
426 .....LeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSerGl 439
      |||... |||...|||
19495 GTCCGCGGACTGGTATCATCCGCTGATGACCGGGCGCATGCTCTCCA 19446
      |||... |||...|||
439 yGlyPheGlnAlaAlaTyrValTyrThrProGluValTyrProThrAlaT 456
      |||... |||...|||
19445 GACGATCTCGCCCAAGATCAAGAAGTGGACCGCTACAGAAGCGCG 19396
      |||... |||...|||
456 hrArgAlaLeuGlyLeuGlyThrCysSerGly.....MetAla 468
      |||... |||...|||
19395 CG....ATCGTGGCGCTCGGCTCGATGGCGGAGCAGCTGCTGCTCTCC 19349
      |||... |||...|||
469 ArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGlu... 484
      |||... |||...|||
19348 GCCCGGGCGCG...GACACCCCGCTCGCGGTGATCGTGTGATCGCGGC 19302
      |||... |||...|||
485 .....SerSerValTyrLeuThrLeuAlaVal 493
      |||... |||...|||
19301 CTGGGTGGGCTTCGGCATCGGCGCTCTCCAGACCGCTACACCTGGCGA 19252
      |||... |||...|||
493 alTyrSer 495
      |||... |||...|||
19251 TCCAGAGC 19244
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-474-933-1

seq_documentation_block:

; Sequence 1, Application US/08474933

; Patent No. 5866410

; GENERAL INFORMATION:

; APPLICANT: Ryan, Michael J.

; APPLICANT: Lotvin, Jason A.

; APPLICANT: Strathyn, Nancy

; APPLICANT: Fantini, Susan E.

; TITLE OF INVENTION: Cloning of the biosynthetic pathway for

; TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmids

; TITLE OF INVENTION: useful therein

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: American Cyanamid Company

; STREET: One Cyanamid Plaza

```
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1
```

alignment_scores:
Quality: 151.50 Length: 536
Ratio: 0.613 Gaps: 27
Percent Similarity: 46.082 Percent Identity: 22.575

alignment_block:

US-09-911-667A-2 x US-08-474-933-1/rev ..

Align seg 1/1 to reverse of: US-08-474-933-1 from: 1 to: 30001

```
18 ArgThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH1 34
|||...|||...|||...|||
20560 CGGGCGGGATTTCGCAAGAACT...TGTCAATCCGCCACCGTAGCCCA 20514
34 sGluValGlnIleGluGlyVal.....HisValGlyL 45
|...|||...|||...|||
20513 CTACGGCTCAGTTCGACAGACCTTGTGGAATTCGGTGGGAGAAATGGGA 20464
45 euGluAlaValGluLeuAspAspGlyAlaAlaValProLysGluPheAla 61
|||...|||...|||...|||
20463 TGGCGAACCACCTCGCAGACCGCGGAGCGCTCGCGGAGGAGCGCGC 20414
62 AsnProThrAspThrPheMetValGluAspAlaValGluAlaIleG1 78
|||...|||...|||...|||
20413 GGCCCCCGCC.....GG 20403
78 yPheGlyLysPheGlnTrpLysLeuSerValLeuThrGly.....LeuA 93
|||...|||...|||...|||
20402 CTTCACCTACCGCCAG...ATCATCACCCTGTCGCCGGCTCCTGTGG 20356
93 laTrpMetAlaAspAlaMetGluMetIleLeuSer..... 105
|||...|||...|||...|||
20355 CGGTGCTGTCGCCGACCTCGACACGACGATGCTCCACCGCGCTCGGT 20306
106 IleLeuAlaProGlnLeuHisCysGluTrpArgLeuProSerTrpGlnVa 122
|||...|||...|||...|||
20305 ACGATCGGCGACCACTGCACGCGCCAG...ACCGTCCAGGCGCTGG.... 20264
122 lAlaLeuLeuThrSerValValPheValGlyMetSerSerThrL 139
|||...|||...|||...|||
```


alignment_scores:

Quality: 147.00 Length: 462
 Ratio: 0.662 Gaps: 17
 Percent Similarity: 48.052 Percent Identity: 21.212

alignment_block:

US-09-911-667A-2 x US-08-928-692-9

Align seg 1/1 to: US-08-928-692-9 from: 1 to: 3000

```

118 ProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMetMe 134
||||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
788 CCATCC...CAATTGGCGCTGCTCTCTATATACACCTTGGGGCTT 834
134 tSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLysT 151
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
835 GCTAGGGCTCTCTGGCAGTCCAGTTTCCACCAAGCATGGCGCTTGT 884
151 hrGlyLeuLysIleSerValLeuTrpThrLeuTyrGlyIleLeuSer 167
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
885 TCACACTGGGAGCACCACCATCTCTTCATCTAGGCCCTATAGCAGAA 934
168 AlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuVal 184
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
935 ACATTGGCGCCAGTATACCGGTATTGAGTATGGTAGGCTTTTATCTGG 984
184 yPheGlyIleGlyValProGlnSerValThrLeuTyr...AlaGluP 200
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
985 TGTGGTGGCGGCGCTCTATCGTGGTGGTCCGATATATATCTCTGAGA 1034
200 heLeuPrometLysAlaArgAlaLysCysIleLeuLeuIleGluValPhe 216
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1035 TTGCTCTCTCTAGTCTAAGGGTCTTTTCGGCGCTTTTACCAATCATG 1084
217 TrpAlaIleGlyThrValPheGluValValLeuAlaValPheValMetPr 233
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1085 ACTAATGTGCGTATTCTTGTGACACAGTCCCTTGTGTACTTCTTGTAGTAA 1134
233 oSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuL 250
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1135 AGGAGTATGGAGATTATCTTGGCAATTCCTGGCGCGATCGGATGCC 1184
250 euPheAlaValLeuCysPheTrpLeuProGluSerAlaArgTyrAspVal 266
|| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1185 TTGAGCTTCTGGGCTCTCTTCTTAGTCCAGAAAGCCCATCTGG..... 1228
267 LeuSerGlyAsnGlnGluLysAlaIleAla.....ThrLeuLysArgIl 281
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1229 CTTGCAGATCACCAGAAAGGAATGTGGCTAGACAGGTGTACACGTAT 1278
281 e..... 281
1279 ACGGGGAGGATCCACACATCGAGCCAGAGGTGGAAGGCTGGAGACAT 1328
282 .....AlaThrGluAsnGlyAlaProMetPro 290
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1329 CTGCGGCGCTGAAACACAGCTCTGGGGAAGACAGTCCCTACTATCACCC 1378
291 LeuGlyLysLeuIleIleSerArgGlnGluAspArgGlyLysMetArgAs 307
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1379 CCATCTGGAATATGCCACCCAGCAACCTCCGGTTTACCATGATGGCAGC 1428
307 pLeuPheThrProHisPheArgTrpThrLeuLeuLeuTrpPheIleT 324
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1429 TATTACTGATTCTTTTACCGCCCTGCCATCATTTGCAGTGGTCGGAGTCA 1478
324 rpPheSerAsnAlaPheSer.....TyrTyrGlyLeu 334
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1479 TGGTTTCCAGCAGTCTCACTGGTGTCAACACGATCATCATGTACAGCGTT 1528
335 ValLeuLeuThrThrGluLeuPheGlnAlaGlyAspValCysGlyIleSe 351

```

```

1529 TCCCTCTTACAGACCATCCTTCCCACCACTGCAGCCCTG.....TTGTC 1572
||||| ||| ||| :||| :||| :||| :||| :||| :||| :|||
351 rSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysGluTyrL 368
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1573 GGTGATCATCTCGGCTCAATCTTGTATCACTCTGCGCTGC..... 1615
368 euSerGluGluAspTyrMetAspLeuLeuTrpThrLeuSerGluPhe 384
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1616 .....TCACCACTA 1624
385 ProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysL 401
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1625 CCT.....GATAAGATTGGTAGAGCTC 1647
401 sThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuP 418
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1648 CTGCTGCTTCTTAAGTATCAGCGCATGGTCTTAATTCGCTCTACTGG 1697
418 heIleCysValGlyArgAsnVal.....LeuThrLeuLeu 429
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1698 CGCTAGCCATCTACTTCAACCTGAAAGCCTTATCGCCCATAGCACTTCTA 1747
430 LeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaIaTyrVa 446
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1748 CTTTTCGTTGCT...TCTTTCGCCCGCTCTAGGCCAGTCCCTTCAT 1794
446 lTyrThrProGlu...ValTyrProThrAlaThrArgAlaLeuGlyLeuG 462
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1795 TTTAGCCCTCTGAACCTGTTGGCCCGAGGCTGTC..... 1828
462 lYThrCysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIle 478
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1829 .....GGCGCCACAGACCTGGCGCTGGAGCGAAGCTGATT 1867
479 AlaGlnValMetLeu..... 483
1868 GCCACGTTTCATCGTGGCACAAATTTTTCGATGTTAAACGATTTTGTGGG 1917
484 ....GluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysL 499
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1918 CGGACGAGGCAAGATCTACTGATCTTTCACGCGATGGCC.....TGTTC 1961
499 euLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGly 515
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1962 TCCTCGGAAGTTTTCATCTACTTGGTGGTGCCG...GAGACCAAGGGG... 2005
516 GlyLeuGlnGluSerSerHisArgGluTrpGlyGln 527
: :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2006 .....AAGGCTAAGCGCGACGAAGTTTGGGAAGG 2035

```

OM of: US-09-911-667A-2 to: EST.* out_format : pfs

Date: Mar 8, 2002 3:01 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODE=frame+g2n.model -DEV-xlp
-O/-cgn2.1/USPTO_spool/US09911667/runat_07032002_130318_29527/app_query.fasta.1.615
-DB=EST -QFMT=fastap -SUFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LJOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000
-QGAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALICEN=15 -MODE=LOCAL -OUTFMT=pfs -NORM-ext -MINLEN=0
-MAXLEN=200000000 -USER=US09911667 -ECGNT_1.4092 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-911-667A-2

Query length: 548

Database: EST.*

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1524.690000

score_list:	Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_est1:AL523335	- 1445.00	2334.61	7.1e-121	972	1	AL523335 LTI_NFL003_N
gb_est1:AL523336	+ 1256.50	2027.56	9.0e-104	1022	1	AL523336 LTI_NFL003_N
gb_est2:BT341271	+ 986.00	1593.42	1.4e-79	570	1	BT341271 368657 MARC 2P1G Sus S
gb_est2:BG703339	+ 944.00	1524.00	1.0e-75	638	1	BG703339 60268528661 NTH_MGC_95
gb_est2:BT394002	+ 939.00	1515.51	3.0e-75	662	1	BT394002 pppin.pk013.c4 Normali
gb_est1:AL5234002	+ 934.00	1505.78	1.0e-74	778	1	AL5234002 AL529154 LTI_NFL001_N
gb_est1:AL523337	+ 898.00	1449.13	1.5e-71	642	1	AL523337 AL563357 LTI_NFL001_N
gb_est1:BF775076	+ 809.00	1307.51	1.2e-63	648	1	BF775076 285128 MARC 3BOV Bos t
gb_est1:AW173250	+ 760.00	1227.43	3.3e-59	487	1	AW173250 xj85b11.xl Soares_NFL
gb_est1:AW663941	- 743.00	1198.10	1.4e-57	578	1	AW663941 hi73h10.xl Soares_NFL
gb_est1:AU220193	+ 619.00	993.98	3.4e-46	742	1	AU220193 AU220193 unpublished C
gb_est1:AA349847	+ 609.00	984.55	1.1e-45	371	1	AA349847 ESF57001 Infant brain
gb_gss:CN5051V9	- 587.00	938.54	4.1e-43	1047	1	AA342778 Tetraodon nigroviridis
gb_gss:CN504275	+ 538.50	861.52	8.0e-39	868	1	AL313898 Tetraodon nigroviridis
gb_est1:AW345918	+ 538.00	868.93	3.1e-39	377	1	AW345918 15969 MARC 4BOV Bos t
gb_gss:CN501UGB	- 512.00	817.07	2.4e-36	996	1	AL167780 Tetraodon nigroviridis
gb_est2:BT296940	- 498.00	802.13	1.6e-35	450	1	BT296940 UI-R-DK0-cqj-d-01-0-U
gb_est1:AW418091	+ 494.00	800.00	2.1e-35	289	1	AW418091 55608 MARC 1BOV Bos t
gb_est1:AW965645	+ 482.00	774.39	5.7e-34	536	1	AW965645 ESF377718 MAGE resequ
gb_gss:CN505PC2	- 478.00	761.28	3.1e-33	1047	1	AL347771 Tetraodon nigroviridis
gb_est1:AA350545	+ 447.50	721.12	5.3e-31	402	1	AA350545 EST57879 Infant brain
gb_est1:BE590388	+ 447.50	715.85	1.0e-30	686	1	BE590388 SB128 Sugar Beet germ
gb_est2:BF412623	- 440.00	708.49	2.7e-30	420	1	BF412623 UI-R-BT1-bnq-a-05-0-U
gb_gss:CN505D2L	- 432.00	686.21	4.7e-29	107	1	AL331878 Tetraodon nigroviridis
gb_est2:F08438	+ 431.00	698.66	9.4e-30	258	1	F08438 HSCZUB081 normalized ind
gb_est1:BE097508	+ 427.00	687.40	4.0e-29	418	1	BE097508 UI-R-B01-apy-c-11-0-U
gb_gss:CN501MWT	+ 417.00	664.23	7.8e-28	842	1	AL151486 Anopheles gambiae GSS
gb_est1:A0067579	+ 407.00	646.87	7.2e-27	941	1	A0067579 A0067579 Sugano mouse
gb_est2:T10079	+ 387.00	625.65	1.1e-25	299	1	T10079 seg1.96 b4HB3MA Cot8-HA
gb_est2:T08271	+ 385.00	620.40	2.2e-25	366	1	T08271 EST06162 Infant Brain, E
gb_est1:AL372624	+ 383.00	617.45	3.1e-25	355	1	AL372624 EST175519 Infant brain
gb_gss:CN5043JM	- 367.00	581.78	3.1e-23	945	1	AL273875 Tetraodon nigroviridis
gb_est1:AL372624	- 358.50	567.18	2.0e-22	1022	1	AL333922 Tetraodon nigroviridis
gb_gss:CN50427T	+ 347.50	555.97	8.4e-22	519	1	AW097386 rs43c02.yl Sommer Pris
gb_est1:AW097386	+ 326.00	526.41	2.1e-20	300	1	CS7850 CS7850 Yuji Kohara unpub
gb_gss:CN505D6L	+ 314.00	495.00	1.1e-18	1003	1	AL332022 Tetraodon nigroviridis
gb_est1:A0207190	+ 308.50	490.73	3.6e-18	624	1	A0207190 A0207190 unpublished C
gb_est1:AT388711	+ 308.00	489.46	4.2e-18	654	1	AT388711 GH19693.5prine GH Dros
gb_est1:AT541594	- 306.00	486.00	6.6e-18	668	1	AL541594 SD02770.5prime SD Dros
gb_gss:CN5051H2	- 302.50	477.17	2.1e-17	918	1	AL316847 Tetraodon nigroviridis
gb_est2:BF641087	+ 299.00	476.30	2.3e-17	563	1	BF641087 NF032A09INIF1068 Insed

gb_gss:CN504Z93 + 288.50 453.05 4.5e-16 1053 ! AL313968 Tetraodon nigrovir
gb_est2:BG591380 + 286.00 453.80 4.1e-16 646 ! BG591380 EST499222 P. infest
gb_est2:BI102569 + 275.50 432.10 6.6e-15 1033 ! BI102569 602889574FI NC1_CG
gb_est1:BE216528 + 274.50 431.53 7.1e-15 928 ! BE216528 HV_CBD0010119f Hord

seq_name: gb_est1:AL523335

seq_documentation_block:
LOCUS AL523335 972 bp mRNA EST 13-FEB-2001
DEFINITION AL523335 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001N02 3
prime, mRNA sequence.
ACCESSION AL523335
VERSION AL523335.1 GI:12786828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 972)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1..972
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001N02"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab.host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com*
BASE COUNT 236 a 283 c 252 g 194 t 7 others
ORIGIN

alignment_scores:
Quality: 1445.00 Length: 292
Ratio: 5.088 Gaps: 0
Percent Similarity: 97.260 Percent Identity: 96.575
alignment_block:
US-09-911-667A-2 x AL523335/rev ..
Align seg 1/1 to reverse of: AL523335 from: 1 to: 972

257 TrpLeuProGluSerAlaTgTyrAspValLeuSerGlyAsnGlnGluLy 273
|||||
970 TGGCTGCTCAAGTGAAGTATGRTGTCTGTCTGTCAGGAACAGGAAA 921
273 sAtaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetp 290
| |||||
920 GCAATCGCCACCTAAAGAGATAGCAACTGAAACGGGCTCCCATGC 872
290 roLeuGlyLysLeuIleSerArgGlnGluAspArgGlyLysMetArg 306
|||||
871 CGCTGGGAAACTCATCTCCACAGAGAGAGACCGAGGCAATGAGG 822
307 AspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrpPheIl 323
|||||

```
821 GACCTTCTCACACCCCATTTTACATGGACAACTTTCCTCTCTGTGGTTAT 772
323 eTrpPheSerAsnAlaPheSerTyrGlyLeuValLeuLeuThrThrG 340
771 ATGGTTTCCATGCATTCCTTACTACGGGTAGTTCTACTCACACAG 722
340 LuLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAla 356
721 MACCTCTTCAGGACGAGAGATGCTCGGGCTCTCCAGTCGGAAGAGGCT 672
357 ValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTyr 373
671 GTAGAGGCAAAATGCAGCCTGGCTCGGAGTACCTGAGTGAGGAGATSA 622
373 rMetAspLeuLeuTyrThrLeuSerGluPheProGlyValLeuValT 390
621 CATGGACTTTCCTGACACACCTCTCTCAGTTTCCAGGTGCTTGTGA 572
390 hrLeuTrpIleLeuAspArgLeuGlyArgLysLysThrMetAlaLeuCys 406
571 CTCGTGGATTAATGACCCCTGGGGCGCAAGAACCATGGCCCTGTGC 522
407 PheValIlePheSerPheCysSerLeuLeuLeuPheIleCysValGlyAr 423
521 TTGTGATCTCTCTCTTCGACCTCTCTGCTGTTTATCTGTGTGGAG 472
423 GAsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyG 440
471 AAATGTGCTTACTCTCTTACTCTTCATTCGAAAGCGTTTATTTCTGGAG 422
440 LyPheGlnAlaAlaTyrValTyrProGluValTyrProThrAlaThr 456
421 GCTTTCAAGCGCATATGTTTACACACCTCGAGGCTTACCCACGGCAAG 372
457 ArgAlaLeuGlyLeuGlyThrCysSerGlyMetAlaArgValClyAlaLe 473
371 CGGGCCCTCGGCTGGCACCTGCAGCGCATGGCAAGAGTGGTCTCT 322
473 uIleThrProPheIleAlaGlnValMetLeuGluSerSerValTyrLeuT 490
321 CANTACTCGCTTCATCGCCAGGTGATGCTGGAATCTCTGTGTACCTGA 272
490 hrLeuAlaValTyrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCys 506
271 CTCTGCAGTTTACAGTGGCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCT 222
507 PheLeuProIleGluThrLysGlyGlyLeuGlnGluSerSerHisar 523
221 TTTTTCCTCATTTGAGACCAAGGCGGAGGACTGCAGGAGTCCAGCCACCG 172
523 gGluTrpGlyGlnGluMetValGlyArgGlyMetHisGlyAlaGlyValT 540
171 GGAGTGGGGCCAGGAGATGCTCGGGGAGGATGCAGGTCAGGTGTGA 122
```

seq_name: gb_est1:AL523336

seq_documentation_block:

LOCUS AL523336 1022 bp mRNA EST 13-FEB-2001
DEFINITION AL523336 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YN02 5
prime, mRNA sequence.

ACCESSION AL523336

VERSION AL523336.1 GI:12786829

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1022)

AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source

1..1022
/location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YN02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 180 a 268 c 326 g 243 t 5 others
ORIGIN

alignment_scores:

Quality: 1256.50 Length: 258

Ratio: 5.006 Gaps: 1

Percent Similarity: 97.287 Percent Identity: 96.899

alignment_block:

US-09-911-667A-2 x AL523336 ..

Align seg 1/1 to: AL523336 from: 1 to: 1022

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValLysPheAr 17
|||||
231 ATGGAGGAGGACTTATTCAGCTAAGCAGCTGCCGGTTGTGAATTCG 280
17 gargThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
|||||
281 TCGCACAGCGAGAGTGCAGAGTTCAGAGGACGACACGGCTTCAGGAGAG 330
34 isGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
|||||
331 ATGAAGTCCAGATTGAWGGGTCCACCTGGGCTAGAGGCTGTGGAGCTG 380
51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
|||||
381 GATGATGGGCAGCTGTGCCAAGAGATTGGCAATCCACCGCATGATAC 430
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGlnT 84
|||||
431 TTTTCATGGTGGAGATGTCAGTGGAGCCATTGGCTTTGGAAATTTTCAGT 480
84 rPLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGlu 100
|||||
481 CGAAGCTGCTCTCTCTACTGCTTGGCTTGGCTGATGCTGATGCCATGGAG 530
101 MetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgLe 117
|||||
531 ATGATGATCTCTCAGCATCTGGCACACACAGTGCATTGGAGTGGAGGCT 579
117 uProSerTrpGlnValAlaLeuLeuThrSerValPheValGlyMetM 134
|||||
580 .CCAAAGTGGCAGGAGTGGCATTCCTGACCTCGGTGGTCTTTGTAGGCATGA 629
134 eTSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLys 150
|||||

```

630 TGTCCAGCTCCAGGCTCTGGGAAATATCTCAGACCAGTACGGCAGGAAA 679
151 ThrGlyLeuLysIleSerValLeuThrTrpThrLeuTyrGlyIleLeuSe 167
|||||
680 ACAGGGCTGAAGATCAGCGTCTGGACACTGTACTATGGCATCCCTAG 729
167 rAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuValG 184
|||||
730 TGCATTTCGCCCGCTGTATAGCTGGATCTTGGTCTCCCGGGCGCTGGTG 779
184 lypheGlyIleGlyValProGlnSerValThrLeuTyrAlaGluPhe 200
|||||
780 GCTTCGGGATCGGAGGATTCGCCAGTCGGTGACGCTGTATGCCGAGTTC 829
201 LeuProMetLysAlaArgAlaLysCysIleLeuLeuIleGluValPheTr 217
|||||
830 CTTCCTCCATGAWGCCAGAGCTAATGATTTGCTGATGAGGTATTCG 879
217 pAlaIleGlyThrValPheGluValValLeuAlaValPheValMetPro 234
|||||
880 GGCCATCGGACAGTGTTCGAGTCGTCCTGGCTGTGTTCTGTGATGCCA 929
234 erLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaValProLeuLeuLeu 250
|||||
930 GCCTGGGCTGGCTGTGGCTGCTCATC...TCTCAGCTGTCGCCCTCCCTC 976
251 PheAlaValLeuCysPheTrpLeu 258
|||||
977 TTTCGCCGTGCTGTGTTCTGGCTG 1000

```

seq_name: gb_est2:B1341271

```

seq_documentation_block:
LOCUS B1341271 570 bp mRNA EST 30-JUL-2001
DEFINITION 368657 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION B1341271
VERSION B1341271.1 GI:15034560
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 570)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,F.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCAGG
Plate: 108 row: D column: 4
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..570
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

```

FEATURES
source

seq_name: gb_est2:BG703339

```

seq_documentation_block:
LOCUS BG703339 638 bp mRNA EST 07-MAY-2001
DEFINITION 602685286F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4817829 5',
mRNA sequence.
ACCESSION BG703339
VERSION BG703339.1 GI:13975571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

BASE COUNT 114 a 147 c 156 g 153 t
ORIGIN
alignment_scores:
Quality: 986.00 Length: 190
Ratio: 5.189 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.895
alignment_block:
US-09-911-667A-2 x B1341271 ..
Align seg 1/1 to: B1341271 from: 1 to: 570

```

```

207 AlalysCysIleLeuLeuLeuIleGluValPheTrpAlaIleGlyThrValph 223
|||||
1 GCTAAGTGCATTTTGGCTGATTGAGGTGTCTGGGCCATCGGGACGGTGT 50
223 eGluValValLeuAlaValPheValMetProSerLeuGlyTrpArgTrpL 240
|||||
51 CGAGGTCTATCCTGGCTGTGTGTGTGATGCCAGCCTGGGCTGGCGTTGGC 100
240 euLeuIleLeuSerAlaValProLeuLeuLeuPheAlaValLeuCysPhe 256
|||||
101 TGCTCCTCTCTTTCAGCTGTCCCGCTCCTCTCTCTCTCTCTCTCTCTCT 150
257 TrpLeuProGluSerAlaArgTyrAspValLeuSerGlyAsnGlnGluLy 273
|||||
151 TGGCTGCCAGAGAGTGCAGATATGATGTCTCTCTGGAAACACAGGAAAA 200
273 sAlaIleAlaThrLeuLysArgIleAlaThrGluAsnGlyAlaProMetP 290
|||||
201 GGCCATTGCCACCTTAAAGAGGATAGCCACGAAACGGAGCCGCCCATGC 250
290 roLeuGlyLysLeuIleSerArgGlnGluAspArgGlyLysMetArg 306
|||||
251 CTCTGGGGAAGCTCATCATCTCCAGACAGGAAGACCGAGGCAAAATGAGG 300
307 AspLeuPheThrProHisPheArgTrpThrThrLeuLeuLeuTrpPheIl 323
|||||
301 GACCTTTTTCACACCCCTTTTAGATGGACAACCTTTTGTGTGTGTAT 350
323 eTrpPheSerAsnAlaPheSerTyrGlyLeuValLeuLeuThrThrG 340
|||||
351 ATGGTTTTCCAATGCGTTTCTTACTATGGACTAGTGTCTGCTCACCCTG 400
340 luLeuPheGlnAlaGlyAspValCysGlyIleSerSerArgLysLysAla 356
|||||
401 AGCTCTTCCAGGCTGGAGATCTCTGCAGCATCTCCAGCCGGAAGAGGCC 450
357 ValGluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTy 373
|||||
451 GTAGAGGCAAAATGCAGCCTGGCGCTGCAGTACCTGAGTAAGGAGGATTA 500
373 rMetAspLeuLeuTrpThrThrLeuSerGluPheProGlyValLeuValT 390
|||||
501 CATGGACCTCTGTGGACCAAGCTCTCTGAGTTTCCAGGTGCTCTTGTGA 550
390 hrLeuTrpIlelleAspArg 396
|||||
551 CTCTCTGGATTATCGACCGC 570

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 638)
NIH-MGC http://mqc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10718 row: 1 column: 22
High quality sequence stop: 638.

FEATURES

source
1..638
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817829"
/clone_lib="NIH-MGC.95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NIH), National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 133 a 149 c 212 g 144 t
ORIGIN

alignment_scores:

Quality: 944.00 Length: 208
Ratio: 4.697 Gaps: 5
Percent Similarity: 96.635 Percent Identity: 95.192

alignment_block:

US-09-911-667A-2 x BG703339 ..
Align seg 1/1 to: BG703339 from: 1 to: 638
1 MetGluGluAspLeuPheGlnLeuArgGlnLeuProValValTysPheAr 17
25 ATGGAGGAGGACTTATTCAGCTAAGGCAGCTGCCGGTTGTCAATTCGG 74
17 gArgThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGluH 34
75 TCGCACAGCGGAGAGTGAAGTCAAGGTCAGAGGACACACGGCTTCAGGAGC 124
34 IsGluValcInleGluGlyValHisValGlyLeuGluAlaValGluLeu 50
125 ATGAAGTCCAGATTGAAGGGTCCACGTGGGCTTACAGCGCTGTGAGCTG 174
51 AspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspAspTh 67
175 GATGATGGGCGAGCTGTGCCCAAGGAGTTTGCCAAATCCCACTGATGATAC 224
67 rPheMetValGluAspAlaValGluAlaIleGlyPheGly.LysPheGln 83
225 TTTCATGGTGAAGATGACGTGAAGGACCAATGGCTTTGGAAACAATTTCAG 274
84 TrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetGl 100
275 TGGAGAGTGTCTGTCTACCTACCTGGCTTGGCTGGATGGCTGATGCATGGA 324

100 uMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgL 117
|||||
325 GATGATGATCCTCAGCATCTGGCACACAGCTGCATTCGAGTGGAGGC 374
117 euProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMet 133
|||||
375 TCCCAAGCTGGCAGGTGGCATTCCTGACCTCGCTGCTTTGTAGGCATG 424
134 MetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArg.L 150
|||||
425 ATGTCCAGCTCCACGCTCTGGGGAATATCTCAGACCAGTACGCGAGGA 474
150 ysThrGlyLeuLysIleSerValLeuTrpThrLeuTyr.TyrGlyIleLe 166
|||||
475 AAACAGGCTGACGATCAGCTGCTGGACTCTGTACCTATGGCATCCT 524
166 uSerAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuV 183
|||||
525 TAGTGCATATGCGCCGTG.TATAGCTGGATCCTGGTCTCCGGGGCTGG 573
183 aiClyPheGlyIle.GlyGlyValProGln.SerValThrLeuTyrAlaG 199
|||||
574 TGGGCTTCGGATACGAGGAGGTTCCTCCCAAGTCGGTGACGCTGATGCCG 623
199 luPheLeuProMet 203
|||||
624 AGTACTTCCCATG 637
seq_name: gb_est2:BI394002
seq_documentation_block: 662 bp mRNA EST 06-AUG-2001
LOCUS BI394002
DEFINITION pgpln.pk013.c4 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus cDNA clone pgpln.pk013.c4 5' similar to
gbIAAC78627.1 (AF060173) SV2 related protein [Rattus norvegicus],
mRNA sequence.
ACCESSION BI394002
VERSION BI394002.1 GI:15087284
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 662)
Porter, T.E. and Cogburn, L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFAFS Animal Genome Project
Unpublished (2001)
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.
FEATURES
Location/Qualifiers
1..662
/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pgpln.pk013.c4"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5,
w7,w9)"
/lab_host="E. Coli BMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"
BASE COUNT 160 a 145 c 169 g 177 t 11 others
ORIGIN

```
alignment_scores:
  Quality: 939.00      Length: 199
  Ratio: 4.968        Gaps: 0
Percent Similarity: 95.980 Percent Identity: 90.955

alignment_block:
US-09-911-667A-2 x B1394002 ..

Align seg 1/1 to: B1394002 from: 1 to: 662

350 IieSerSerArgLysLysAlaValGluAlaLysCysSerLeuAlaCysG1 366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6 ATTTCCAGTAGAAGAAAGAAAGCTTAAAGCAAGTGCAGCGCTGACCTGCA 55

366 uTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrLeuSerG 383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
56 GTATCTGCAGAGAGAGACTACACTGATCTGCTGCGACAAAGCTGTCAG 105

383 IuPheProGlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArg 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 AATTTCCAGGTGTGTAGTAACACTGTGGATTATTGATCGGATAGCCCG 155

400 LysLysThrMetAlaLeuCysPheValIlePheSerPheCysSerLeuLe 416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 AAGAAACCATGCGCTGCTCTTCTTCTTCATTTTGCAGCGCTGCT 205

416 uLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPheIleA 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 GCTGTTTCTCTGTGTGAAGAAATGTTTACTGTGCTGCTCTTCATTTG 255

433 IaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrThrPro 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
256 CAAGAGCTTTTATTTCAGAGGATTTCAGGCTGCTTATGTTTACACTCCG 305

450 GluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerG1 466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
306 GAGGTTTACCAACAGCCAGCGCTGCTGCGCTGGGAACATGCAGTGG 355

466 yMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnValMetL 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
356 AATGGCCAGAGTGGAGCCCTCATACACCATTCATTGCACAGGTGATGT 405

483 euGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeu 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
406 TGAATCTTCAGTCTATTAAACACTGGCAGTTACAGTGGATTGCGCTG 455

500 LeuAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyG1 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
456 CTGGCTGCTGTGGCTTCTCTTCTTCTTCCCATTTGAACAAAGGTCGTGG 505

516 yLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGlyArgG 533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
506 CTTCAGGAGTCCAGCCACAGAGATGGGACAGAGATGGTTGGGAGAG 555

533 lyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGlnGlu 548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
556 GATCTCATTCGCCAGAGTCCACAGGTCTAAGTCTGCGGTGCACAGGAG 602
```

seq_name: gb_est1:AL529154

```
seq_documentation_block:
LOCUS      AL529154      778 bp      mRNA      EST      13-FEB-2001
DEFINITION AL529154 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 5
prime, mRNA sequence.
ACCESSION  AL529154
VERSION    AL529154.1 GI:12792647
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
1 (bases 1 to 778)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
            source
            1..778
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DD004YD17"
            /clone_lib="LTI_NFL001_NBC4"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by Life Technologies. Contact : Feng Liang Life
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"

BASE COUNT      150 a      193 c      254 g      166 t      15 others
ORIGIN

alignment_scores:
  Quality: 934.00      Length: 199
  Ratio: 4.968        Gaps: 1
Percent Similarity: 94.472 Percent Identity: 93.467

alignment_block:
US-09-911-667A-2 x AL529154 ..

Align seg 1/1 to: AL529154 from: 1 to: 778

1 MetGluGluAspLeuPheGlnLeuArgGlnLeuPro.ValValLysPheA 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
182 ATGGAGGAGGACTTATTCAGCTAAGCAGCGCCGCGGTTGTCAAATCC 231

17 rGArgThrGlyGluSerAlaArgSerGluAspThrAlaSerGlyGlu 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
232 GTCCGACAGCGGAGAGTGCAGGTCAGAGGACGACACGCGCTTCAGGAG 281

34 HisGluValGlnIleGluGlyValHisValGlyLeuGluAlaValGluLe 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
282 CATGAAGTCCAGATTGAAGGGGTCCACGTGGGCTTAGAGGCTGTGGAGCT 331

50 uAspAspGlyAlaAlaValProLysGluPheAlaAsnProThrAspSpt 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 GCATGATGGGCGAGCTGTGCCCAAGGAGTTTGCCAAATCCACCGATGATA 381

67 hrPheMetValGluAspAlaValGluAlaIleGlyPheGlyLysPheGln 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
382 CTTTCATGTTGGAGAGTGCAGTGGGAAGCCATTGGCTTTGGAAAATTTTCAG 431

84 TrpLysLeuSerValLeuThrGlyLeuAlaTrpMetAlaAspAlaMetG1 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
432 TGGAAAGTGTCTGTCTCACTGGCTTGGCTTGGATGGCTGATGCCATGGA 481

100 uMetMetIleLeuSerIleLeuAlaProGlnLeuHisCysGluTrpArgL 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
482 GATGATGATCCTCAGCATCTCTGGCACCACAGCTGCTTTCGAGTGGAGGC 531

117 euProSerTrpGlnValAlaLeuLeuThrSerValValPheValGlyMet 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
532 TCCCAAGCTGGCAGGTGATGCTGCTGAMCTCGGTGCTCTTTTATAGGCATG 581
```


134 MetSerSerSerThrLeuTrpGlyAsnIleSerAspGlnTyrGlyArgLy 150
 |||||
 582 AMGTCAGCTCCACGCTCTGGGAAATATYTCAGACCAGTACCGCAGGAA 631
 |||||
 150 sThrGlyLeuLysIleSerValLeuThrTrpThrLeuTyrGlyIleLeuS 167
 |||||
 632 AACARGGCTGAAGATCAGCGTCTCTGGACTCTKTCATGATGCATCCCTTA 681
 |||||
 167 erAlaPheAlaProValTyrSerTrpIleLeuValLeuArgGlyLeuVal 183
 |||||
 682 GTGCATTTGGCGCCCTKWTAGCTGCATCTGGTCTCCGGGCGCTGGTG 731
 |||||
 184 GlyPheGlyIleGlyGlyValProGlnSerValThrLeuTyrAla 198
 |||||
 732 GGCCTGGGATCGGAGGAGTTCGCCAGTCGGTGACGCTKTTTGCC 776
 |||||

seq_name: gb_est1:AL563357

seq_documentation_block:
 LOCUS AL563357 642 bp mRNA EST 16-FEB-2001
 DEFINITION AL563357 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD004YD17 3
 prime, mRNA sequence.
 ACCESSION AL563357
 VERSION AL563357.1 GI:12912671
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source
 1..642
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DD004YD17"
 /clone_lib="LTI_NFL001_NBC4"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
 8371 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 149 a 192 c 170 g 117 t 14 others
 ORIGIN

alignment_scores:
 Quality: 898.00 Length: 202
 Ratio: 4.854 Gaps: 1
 Percent Similarity: 91.584 Percent Identity: 90.099

alignment_block:

US-09-911-667A-2 x AL563357/rev ..

Align seg 1/1 to reverse of: AL563357 from: 1 to: 642

348 CysGlyIleSerSerArgLysLysAlaValGluAlaLysCysSerLeuAl 364
 |||||
 637 TGCAGCTTCTCCAGTCGGAARAAGCTGAAGAGGCAAAATGACGCTGCG 588
 |||||

364 aCysGlyTyrLeuSerGluGluAspTyrMetAspLeuLeuTrpThrL 381
 |||||
 587 CTGCGAGTACCTGAGTGAGGAGGATWACACTGGACTTGTGTGCACCA 538
 |||||
 381 euSerGluPheProGlyValLeuValThrLeuTrpIleLeuAspArgLeu 397
 |||||
 537 TCTCTGAGTTTCAGGTGCTCTGTGACTCTGTGGATTATTGACCGCTG 488
 |||||
 398 GlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCysSe 414
 |||||
 487 GGGCGCAAGAAGACCATGGCCCTGGGCTTTGTGCATCTTCTCTTCNGCAG 438
 |||||
 414 rLeuLeuLeuPheIleCysValGlyArg.AsnValLeuThrLeuLeuLeu 430
 |||||
 437 CCNCCNGCTGTTTNNCTGTGTGGAAGAAATGNGCTCACTCTGTAAC 388
 |||||
 431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTy 447
 |||||
 387 TTTCATGGCAAGAGCGCTTATTCTTGAGGCTTTTCAAGCGGCATATGTTNA 338
 |||||
 447 rThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrC 464
 |||||
 337 CACACCTGAGGTCTACCCACGCAACGGCGCCCTCGGCTGGGCGACCT 288
 |||||
 464 ysSerGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGln 480
 |||||
 287 GCAGCGGCATGCGCAAGAGNGGGCTCTCANTACTCCGTTTCATCGCCAG 238
 |||||
 481 ValMetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCy 497
 |||||
 237 GTGATGCTGGAATCTCTGTGTACCTGACTGTGGCAGTTAACAGTGGCNG 188
 |||||
 497 sCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrLysG 514
 |||||
 187 CTGCTCTCTGGCTGCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 138
 |||||
 514 lylGlyGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetVal 530
 |||||
 137 GCCGAGNACTGCGAGGAGTCCAGCCACCGGAGTGGGGCCAGGAGATGGCTC 88
 |||||
 531 GlyArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGI 547
 |||||
 87 GGGCGAGGAATGACGGTGCAGGTGTTTACAGGTGCGGACTCTGGCTCTCA 38
 |||||

seq_name: gb_est2:BF775076

seq_documentation_block:

LOCUS BF775076 468 bp mRNA EST 25-APR-2001
 DEFINITION 285128 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BF775076

VERSION BF775076.1 GI:12122976

KEYWORDS EST.

SOURCE cow.

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

1 (bases 1 to 468)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL

MEDLINE

COMMENT

21180013

Co.tact: Smith TPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTTCCGAGTCACGACG
Plate: 86 row: E column: 5
Seq primer: ATTAGCTGACACTATAG.

FEATURES

source
1. .468
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH108"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 87 a 126 c 139 g 116 t
ORIGIN

alignment_scores:
Quality: 809.00 Length: 156
Ratio: 5.186 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-911-667A-2 x BF775076 ..
Align seg 1/1 to: BF775076 from: 1 to: 468

163 TyrGlyIleLeuSerAlaPheAlaProValTyrSerTrpIleLeuValLe 179
1 TACGGCATCCCTCAGTGCCTTTGCCCGCGGTACAGCTGGATCCTGGTGT 50
179 uArgGlyLeuValGlyPheGlyIleGlyValProGlnSerValThrL 196
51 CCGGGCCCTGGTGGCTTCGGAGTCGGAGGGTCCCCAGTCGGTCACAC 100
196 euTyrAlaGluPheLeuProMetLysAlaArgAlaLysCysIleLeuLeu 212
101 TGTATGCTGAGTTCCTTCCCATGAAAGCCAGAGCTAAATGATTTTGTG 150
213 IleGluValPheTrpAlaIleGlyThrValPheGluValValLeuAlaVa 229
151 ATTGAGGTGTTCTGGCCCATCGGACAGTGTTCGAGGTGCTCTGGCTGT 200
229 IpheValMetProSerLeuGlyTrpArgTrpLeuLeuIleLeuSerAlaV 246
201 GTTTGTGATGCCAGCTGGCTGGCTGGCTGCTCATCTCTCAGCTG 250
246 alProLeuLeuLeuPheAlaValLeuCysPheTrpLeuProGluSerAla 262
251 TCCCACTCCTCTCTTTGCGGTCTGTGTTTTTGGCTGCCGAGAGTGGC 300
263 ArgTyrAspValLeuSerGlyAsnGlnGluLysAlaIleAlaThrLeuLy 279
301 AGGTATGATGTCTGTCCGGAACACGAGAGAGCCATCGCCACCTTAA 350
279 sArgIleAlaThrGluAsnGlyAlaProMetProLeuGlyLysLeuLeI 296
351 GAGAAATAGCCAGGAAACGAGAGCTCCATGCCCTCTGGGAAACTATCA 400
296 leSerArgGlnGluAspArgGlyLysMetArgAspLeuPheThrProHis 312
401 TTTCCAGACGAAAGACCGAGGCAAAATGAGGGACCTTTTTCACACCCCAT 450

313 PheArgTrpTrpThrLeu 318
|||||
451 TTTAGATGCACAACCTTG 468
seq_name: gb_est1:AW173250
seq_documentation_block:
LOCUS AW173250 487 bp mRNA EST 16-NOV-1999
DEFINITION x185b11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663997 3' similar to TR:Q92217 Q92217 SV2 RELATED PROTEIN.
.: mRNA sequence.

ACCESSION AW173250
VERSION AW173250.1 GI:6439198
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 487)
REFERENCE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cygabs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES

source
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663997"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH108"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 126 a 143 c 133 g 84 t 1 others
ORIGIN

alignment_scores:
Quality: 760.00 Length: 162
Ratio: 4.872 Gaps: 0
Percent Similarity: 96.296 Percent Identity: 90.741

alignment_block:

US-09-911-667A-2 x AW173250/rev ..

Align seg 1/1 to reverse of: AW173250 from: 1 to: 487

386 GlyValLeuValThrLeuTrpIleIleAspArgLeuGlyArgLysLysTh 402
487 GGTGTCATTGTGACTCTGTGGATTATTACCCGATGGGGCGCAAGAC 438
402 rMetAlaLeuCysPheValIlePheSerPheCysSerLeuLeuLeuPheI 419
|||||
437 CATGGCCNTGTGCTGTCATCTTCTCTGTCAGCTCGTGTGTTTA 388
419 leCysValGlyArgAsnValLeuThrLeuLeuPheIleAlaArgAla 435

```

|||||.....:|||||.....
387 TCTGTGTGGAGAGATGCTCTACTCTGTACTCTTCATTGCAAGAGCG 338
436 PheIleSerGlyGlyPheGlnAlaIatYrValTyrThrProGluValTy 452
|||||.....:|||||.....
337 TTTATTATGAGGCTTTCAACGGGCATATGTTTACACACGTGAGGTCTA 288
452 rProThrAlaThrArqAlaLeuGlyLeuGlyThrCysSerGlyMetAlaA 469
|||||.....:|||||.....
287 CCCACGGCAACGGGGCCCTCGGCCTGGCACCTGGCAGCGCATGGCAA 238
469 rGValGlyAlaLeuIleThrProPheIleAlaGlnValMetLeuGluSer 485
|||||.....:|||||.....
237 GAGTGGGTGCTCTCATCTCCGTTTCATCGCCAGGTGCTGGAATCC 188
486 SerValTyrLeuThrLeuAlaValTyrSerGlyCysCysLeuLeuAlaI 502
|||||.....:|||||.....
187 TCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTGCTCTGCTGC 138
502 aLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyLeuGlnG 519
|||||.....:|||||.....
137 CCTGGCTCTCTGCTTTTGGCCATTGAGACCAAGGCCGAGGACTGCAGG 88
519 luSerSerHisArgGluTrpGlyGlnMetValGlyArgGlyMetHis 535
|||||.....:|||||.....
87 AGTCCACCCACGGGAGTGGGGCAGAGATGTCGCCCAAGGAATGCAC 38
536 GlyAlaGlyValThrArgSerAsnSerGlySerGln 547
|||||.....:|||||.....
37 GGTTCAGCTGTTCACAGGTGCAACTCTGGGCTCTCAG 2

```

seq_name: gb_est1:AW663941

seq_documentation_block: 578 bp mRNA EST 06-APR-2000
 LOCUS AW663941
 DEFINITION hi73h10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 IMAGE:2977987 3' similar to TR:Q92217 Q92217 SV2 RELATED PROTEIN.
 ; mRNA sequence.

ACCESSION AW663941 GI:7456480
 VERSION AW663941.1
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 578)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cyapbs-r@mail.nih.gov

This clone is available royalty-free through LINL: contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 464.
 Location/Qualifiers
 1..578

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2977987"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NbHL19W, testis NHT, and B-cell
 NCI_CGAP_GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.W.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo. 111 t 1 others
 BASE COUNT 137 a 170 c 159 g
 ORIGIN

alignment_scores:

Quality: 743.00 Length: 151
 Ratio: 5.020 Gaps: 1
 Percent Similarity: 98.013 Percent Identity: 98.013

alignment_block:

US-09-911-667A-2 x AW663941/rev ..

Align seg 1/1 to reverse of: AW663941 from: 1 to: 578

```

399 ArgLysLysThrMet..AlaLeuCysPheValIlePheSerPheCysSerL 415
|||||.....:|||||.....
576 CGCAAAAGACCATGGGCCCTGTCTTGTTCATCTTCTCTCTGTCAGCC 527
415 euleLeuPheIleCysValGlyArgAsnValLeuThrLeuLeuPhe 431
|||||.....:|||||.....
526 TCCTGCTGTTATCTGTGTGGAGANATGCTCACTCTGTACTCTTCTTAC 477
|||||.....:|||||.....
432 IleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTyrTh 448
|||||.....:|||||.....
476 ATTGCAAGAGCGTTTATTCTGGAGGCTTTCAAGCGGCATATGTTTACAC 427
|||||.....:|||||.....
448 rProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysS 465
|||||.....:|||||.....
426 ACCTGAGGTCTACCCACGCGCAACGCGGCCCTCGGCCCTGGGCACCTGCA 377
465 erGlyMetAlaArgValGlyAlaLeuIleThrProPheIleAlaGlnVal 481
|||||.....:|||||.....
376 GCGGCATGCGCAGAGTGGGTGCTCTCATCTACCTCCGTTATCGCCCAAGTG 327
|||||.....:|||||.....
482 MetLeuGluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysCy 498
|||||.....:|||||.....
326 ATGCTGGAATCTCTGTGTACCTGACTCTGGCAGTTTACAGTGGCTGCTG 277
|||||.....:|||||.....
498 sL..uLeuAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyC 515
|||||.....:|||||.....
276 CCTCTCTGGCTGGCCCTGCGCTCTCTGCTTTTGGCCATTGAGACCAAGGCC 227
|||||.....:|||||.....
515 lGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetValGly 531
|||||.....:|||||.....
226 GAGGACTGCGAGGATCCACGCCAGCGGAGTGGGGCCAGGAGATGTCGGC 177
|||||.....:|||||.....
532 ArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySerGlnG 548
|||||.....:|||||.....
176 CGAGGAATGCGAGGTGCGAGGTGTACCGAGTGGAACTCTGGCTCTCAGGA 127
|||||.....:|||||.....
548 u 548
126 A 126

```

seq_name: gb_est1:AU220193

seq_documentation_block:

LOCUS AU220193 742 bp mRNA EST 17-JUL-2001
 DEFINITION AU220193 unpublished oligo-capped cDNA library, stage L1
 Caenorhabditis elegans cDNA clone yk86g01 3', mRNA sequence.

ACCESSION AU220193 GI:14858350

VERSION AU220193

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 742)

AUTHORS Kohara, Y., Shin-i.T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

TITLE A complementary view of the C.elegans genome

450	uValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyM	467
254	AGTTTATCTCAACATTGAGACCGGTGGACTGGAAACGTGCTCGGA	205
467	eAlaArgVaIGlyAlaLeuIleThrProPheIleAlaInValMetLeu	483
204	TGGCAAGAATTGGAGCAATTGTGACTCCATTATCGCTTCAGGTTGC	155
484	GluSerSerValTyrLeuThrLeuAlaValTyrSerGlyCysLeuLe	500
154	GAGAAATCTCTATCTTACCCTATAGGAATTTATGGAACCTCGGCA	105
500	uaLaAlaLeuAlaSerCysPheLeuProIleGluThrLysGlyGlyL	517
104	TGGATTGTAGTCTTCTCTAGGTTTACCAATTGAGACAAGGACGCTA	55
517	euGlnGluSer	520
54	TGATGGACTCC	44
seq_name: gb_estl:AA349847		
seq_documentation_block:		
LOCUS	AA349847	371 bp mRNA EST 21-APR-1997
DEFINITION	EST57001 Infant brain Homo sapiens cDNA 5' end similar to similar to C. elegans hypothetical protein, cosmid ZK637.1, mRNA sequence	
ACCESSION	AA349847	
VERSION	AA349847.1 GI:2002320	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.	
TITLE	3,400 expressed sequence tags identify diversity of transcripts from human brain	
JOURNAL	Nature Genet. 4, 256-267 (1993)	
MEDLINE	93364420	
COMMENT	Other_ESTs: EST57000 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/ldb/hgi/hgi.html) Seq primer: M3 Reverse.	
FEATURES		
source		
1..371		
/organism="Homo sapiens"		
/db_xref="AFCC (inhost):150872"		
/db_xref="taxon:9606"		
/clone_lib="Infant brain"		
/sex="female"		
/dev_stage="infant"		
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;		
Site_2: NotI"		
BASE COUNT	75 a	90 c 93 g 110 t 3 others
ORIGIN		
alignment_scores:		
Quality:	609.00	Length: 123
Ratio:	5.118	Gaps: 0
Percent Similarity:	96.748	Percent Identity: 96.748
alignment_block:		
US-09-q11-667A-2 v AA349847		

Align seg 1/1 to: AA349847 from: 1 to: 371

```
329 pheSerTyrTyrGlyLeuValLeuLeuThrThrGluLeuPheGlnAlaG1 345
|||||
1 TTCTCTTACTACGGGTAGTTCTACTACACAGAACTCTTCCAGGCAGG 50
|||||
345 yaspValCysGlyIleSerArgLysLysAlaValGluAlaLysCysS 362
|||||
51 AGATGCTCGGGCATCTCCAGTCGGAAGAGGCTCTAGAGGCAAAATGCA 100
|||||
362 erLeuAlaCysGluTyrLeuSerGluGluAspTyrMetAspLeuLeuT 378
|||||
101 GCCTGGCCTGGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGT 150
|||||
379 ThrThrLeuSerGluPheProGlyValLeuValThrLeuTrrPheIle 395
|||||
151 ACCACCCCTCTGAGCTTTCCAGGTCTCTGTTGACTCTNTGATTTATGA 200
|||||
395 pArgLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePheSer 412
|||||
201 CCGCTGGGGCGCAAGAACATGGCCCTGTGCTTTGTATCTTCTCTCT 250
|||||
412 heCysSerLeuLeuLeuPheIleCysValGlyArgAsnValLeuThrLe 428
|||||
251 TCTGAGCCTCTCTGCTGTTTATCTGTGTTGGAAGAAATGCTCACTCT 300
|||||
429 LeuLeuPheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAla 445
|||||
301 TTACTCTTCATTTCAAGAGCGCTTTATTTCTGGAGGCTTTCAAGCGGC 350
|||||
445 rValTyrThrProGluVal 451
|||||
351 TGTTTACAAACTGAGGTC 369
|||||
```

seq_name: gb_gss:CNS05LV9

```
seq_documentation_block:
LOCUS      CNS05LV9      1047 bp      DNA      26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
            053F06 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL343278
VERSION    AL343278.1 GI:8237036
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
            Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 1047)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
            Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
            Unpublished
REFERENCE  2 (bases 1 to 1047)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
            Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
            Unpublished
REFERENCE  3 (bases 1 to 1047)
            Genoscope.
            Direct Submission
            Submitted (12-APR-2000) to the EMBL/GenBank/DBSJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
            Location/Qualifiers

FEATURES
```

source 1. 1047
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="053F06"
/clone_lib="A"
/note="Genoscope sequence ID : C0AA053DC03A1-end : T3"

BASE COUNT 255 a 278 c 269 g 229 t 16 others
ORIGIN

alignment_scores:
Quality: 587.00 Length: 275
Ratio: 3.739 Gaps: 6
Percent Similarity: 57.091 Percent Identity: 49.091
alignment_block:
US-09-911-667A-2 x CNS05LV9/rev ..

Align seg 1/1 to reverse of: CNS05LV9 from: 1 to: 1047

```
358 GluAlaLysCysSerLeuAlaCysGluTyrLeuSerGluGluAspTyrMe 374
|||||
881 GAGCTCCGTTGCAGCTGGAGTGAATATCTGAACTCAGACGACTACAA 832
|||||
374 tAspLeuLeuTrrPThrLeuSerGluPhePro..... 385
|||||
831 GGACCTGCTGGACCACCTTTGCTGAAATTTCCAGGCACGACTACTGTTCC 782
|||||
385 ..... 385
781 ATGTGTCAGACAAACCTTAATTCATCCCTGATGAATTTACTGATGTGGCA 732
|||||
386 .....GlyValLeuValThrLeuTrrPheIleAspAr 396
|||||
731 TTGTTGTTGTTTACGCGAGGACTGCTGCTGACACTGTGGCAATCGATCG 682
|||||
396 gLeuGlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheC 413
|||||
691 GCTGGGAGGACGAGACCATGGGCTTTGCTTTCTTCATCTTCTCCATGT 632
|||||
413 ysSerLeuLeuLeuPheIleCysValGlyArg.AsnVal..... 425
|||||
631 GCATTATTCACACTCTACGGCTGTGTGCGCGGTAAAGATTTCATAGCAGCTG 582
|||||
425 ..... 425
581 GTCAGATTCCAACTGGAGAGCCATGCAGNATACTGCCTTATTCTGCC 532
|||||
426 .....LeuThrLeuLeuLeuPheIleAlaArgAlaPheIleSer 438
|||||
531 GCAGGGCTCCATGACCGCTGTGATATTTCATCGCCAGAGCTTTCATTGCA 482
|||||
439 GlyGlyPheGlnAlaAlaTyrValTyrThrProGlu..Val..... 451
|||||
481 GGAGGATTTTCAGGCCGCTNNNGTTTACACTCCAGAGGTAAAGAGCCAGT 432
|||||
451 ..... 451
431 TAGAATCGAGATGACAGTTCAGGCCCATCGCTGTTGTGCAATCTCTGCA 382
|||||
452 ....TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyM 467
|||||
381 GGTGTATCCAAACGGCAGCAGGGCTTTGGGTCTGGGAACAGCAGCGGAA 332
|||||
467 etAlaArgValGlyAlaLeuIleThrProPheIleAlaGln..... 480
|||||
331 TGGCAAGAGTCGGCGCCTGATCACACCATTTTGTGGCAGAGGTGAGGACG 282
|||||
480 ..... 480
281 TCCTTTGCTTATTGAGCTCAAAACCTCAAAAGATGTGTCGTACACCTGC 232
|||||
481 .....ValMetLeuGluSerSerValTyrLeuThrLeuA 492
```

```

231 CGGCTTGTCTAAAGGTGATGCTGGAGTGTGCTGCTGCTGCTGCTGT 182
|||||
492 laValTyrSerGlyCysCysLeuLeuAlaLeuAlaSerCysPheLeu 508
|||||
181 TTGTGTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 132
|||||
509 ProfileGluThrLysGlyGlyGlyLeuGluSerSerHisArgGluTr 525
|||||
131 CCCATCGAGACAGCGGTGCGGGCGCTGCAGGAGTCCAGCCAGCAGAGTG 82
|||||
525 pGlyGlnGluMetValGlyArg 532
|||||
81 GGGCCAGCAAAATGTGGGCCGC 60
|||||

seq_name: gb_gss:CNS04275

seq_documentation_block:
LOCUS      CNS04275      868 bp      DNA      26-JUL-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
040B23 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL313898
VERSION    AL313898.1 GI:9546782
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 868)
AUTHORS   Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissbach,J.
            Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
            Nat. Genet. 25 (2), 235-238 (2000)
20296633
2 (bases 1 to 868)
AUTHORS   Crolius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissbach,J.
            Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish tetraodon nigroviridis
            Genome Res. 10 (7), 939-949 (2000)
20359837
3 (bases 1 to 868)
AUTHORS   Genoscope.
            Direct Submission
            Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source            1..868
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="040B23"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA040CA12C1-end : T7"
BASE COUNT          174 a 222 c 238 g 217 t 17 others
ORIGIN

alignment_scores:
Quality: 538.50      Length: 273
Ratio: 3.614        Gaps: 5
Percent Similarity: 54.579 Percent Identity: 47.619

alignment_block:
US-09-911-667A-2 x CNS04275
..

```

```

Align seg 1/1 to: CNS04275 from: 1 to: 868

376 LeuLeuTrpThrLeuSerGluPhePro..... 385
|||||
16 CTGCTGTGGACCACTTTGTCTGAATTTCCAGGCACCTACTGTTCCCATGT 65
|||||
385 ..... 385
|||||
66 GTCAGACAAACCTTAATTCATCCGCGATGAATTTACTGATGTCGTCATTGT 115
|||||
386 .....GlyValLeuValThrLeuTrpLleIleAspArgLeu 397
|||||
116 TGTGTTTGTGACGACGACTGCTGTCACACTGTGGCAATCGATCGGNTG 165
|||||
398 GlyArgLysLysThrMetAlaLeuCysPheValIlePheSerPheCysse 414
|||||
166 GGGAGGACGAGACCATGGCTTGTGTTTCTTCATCTTCTCCATGTGCAT 215
|||||
414 rLeuLeuLeuPheLleCysValGlyArg..... 423
|||
216 TATTCCACTCTAAGGCTGTGTGCGGCCGTAAAGATTATCAGCAGCTGGTCA 265
|||||
423 ..... 423
|||||
266 GATTCCAAACTGGAGACCATGACGACATAAATGCTTTATTTCGCCGNAG 315
|||||
424 AsnValLeuThrLeuLeuPheIleAlaArgAlaPheIleSerGlyG1 440
|||
316 GNCATCATGACCGGTGATGATATTCATGCCAGGACGTTTCATTCAGGAGG 365
|||||
440 yPheGlnAlaAlaTyrValTyrThrProGlu..... 450
|||||
366 ATTTAGGCGCGCTACGTTTACACTCCAGAGGTAAGAGCCAGCTTAGAA 415
|||||
451 .....Val 451
|||||
416 TGCAGATGACAGTTTCAAGCCCATCGCTGTTGCGCAATCTCTCCAGGTG 465
|||||
452 TyrProThrAlaThrArgAlaLeuGlyLeuGlyThrCysSerGlyMetAl 468
|||||
466 TATCCAAACGCGACAGGCGCTTTGTTGCTTGGAAACAGCAGCGGAATGGC 515
|||||
468 aArgValGlyAlaLeuLleThrProPheIleAlaGln..... 480
|||||
516 AAGAGTCGCGCCCTGATCACCATTGTTGSCACAGGTGAGGAGCTCCT 565
|||||
480 ..... 480
|||||
566 TTGCTTATTGAGCTCAAAACCTCAAAAGATGTGTCGTACACCTGCCGCG 615
|||||
481 .....ValMetLeuGluSerSerValTyrLeuThrLeuAlaVa 493
|||||
616 CTGTGCTTAAAGGTGATGCTGGAGTCTGCGGTACTGCTGCTGCTGTTGT 665
|||||
493 lTyrSerGlyCysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProI 510
|||||
666 GTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 715
|||||
510 leGluThrLysGlyGlyGlyLeuGlnGluSerSerHisArgGluTrpGly 526
|||||
716 TCGAGACGAC.GGGTCGGCGCCTGCAGAGTCCAGCAGCAGTGGGCG 764
|||||
527 GlnGluMetValGlyArgGlyMetHisGly.....AlaGlyValThrAr 541
|||||
765 CAGGAAATNGTGGCGCGCCTCGTCCCGCGGCTTAGCGAGGAGNCCCTCA 814
|||||
541 gSerAsnSerGlySerGln 547
|||
815 TTCTCTCTCGGGTCACAA 833

seq_name: gb_est1:AW345918

```

```

seq_documentation_block: 377 bp mRNA EST 25-APR-2001
LOCUS AW345918
DEFINITION 15969 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW345918
VERSION AW345918.1 GI:6843628
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 377)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.F.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
,G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 8 row: B column: 9
Seq primer: ATTAGGTGACACTATAG.
FEATURES
source
Location/Qualifiers
1..377
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT 62 a 112 c 130 g 73 t
ORIGIN

alignment_scores:
Quality: 538.00 Length: 120
Ratio: 4.719 Gaps: 2
Percent Similarity: 95.000 Percent Identity: 91.667

alignment_block:
US-09-911-667A-2 x AW345918 ..
Align seg 1/1 to: AW345918 from: 1 to: 377

431 PheIleAlaArgAlaPheIleSerGlyGlyPheGlnAlaAlaTyrValTy 447
|||||
3 TTCATTGCGGAGCATGGATTCTTGAGGCTTCCAGCAGCCCTACGTTTA 52

447 rThrProGluValTyrProThrAlaThrArgAlaLeuGlyLeuGlyThrc 464
|||||
53 CACACCTGAGGCTTACCCACGCGCACCAGCGCTGGGCTGGGCACCT 102

464 ysSerGlyMetAlaArgValGlyAlaLeuIleThrPropheIleAlaGln 480
|||||
103 GCAGCGCATGGCGAGGTGGGGCCCTCATCACCCCGTTCATTGCTCAG 152

481 ValMet.LeuGlu.SerSerValTyrLeuThrLeuAlaValTyrSerGly 496
||||| :||| ||||||||||||||||||| |||||||

```

```

153 GTGATCGGTGGAGATCATCGGTCTACCTGACCGCTGCCGGTTACAGTGGC 202
497 CysCysLeuLeuAlaAlaLeuAlaSerCysPheLeuProIleGluThrIy 513
|||||
203 TGTGCTGCTCTGCTGCTGCCCTGCGCTCTCTGCTTTTTCCTTATCGAGACCAA 252
513 sGlyGlyGlyLeuGlnGluSerSerHisArgGluTrpGlyGlnGluMetV 530
|||||
253 AGGCCCGGACTGCAGGAGTCCAGCCACCGGAATGGGGGCGAGAGATGG 302
530 aIGlyArgGlyMetHisGlyAlaGlyValThrArgSerAsnSerGlySer 546
|||||
303 TTGGCCGAGGGGCACACGCGCACAGGTGTGCGCCAGGTGCGAACTCGGGCTCT 352
547 GlnGlu 548
|||||
353 CAGGAG 358

```